OM nucleic - nucleic search, using sw model

```
            GenCore version 5.1.8
    Copyright (c) 1993 - 2006 Biocceleration Ltd.
```

Run on: May 31, 2006, 10:18:03 ; Search time 17609 Seconds
(without alignments)
11497.390 Million cell updates/sec

Title: US-10-501-930-2
Perfect score: 3166
Sequence: 1 cggagcggcgagtcggtgcc.........gctattagcaaaaaaaaaa 3166
Scoring table: IDENTITY.NUC
Gapop 10.0 , Gapext 1.0

Searched: 636136 seqs, 3197371025 residues
Total number of hits satisfying chosen parameters: 12732272

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database : GenEmbl:*
```
1:  gb_env:*
2:  gb_pat:*
3:  gb_ph:*
4:  gb_pl:*
5:  gb_pr:*
6:  gb_ro:*
7:  gb_sts:*
8:  gb_sy:*
9:  gb_un:*
10: gb_vi:*
11: gb_ov:*
12: gb_htg:*
13: gb_in:*
14: gb_om:*
15: gb_ba:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3166	100.0	3166	6	AB059554	AB059554 Mus muscu
c 2	2361	74.6	110000	12	BX322642_0	BX322642 Mus muscu
3	2361	74.6	110000	12	BX322642_2	Continuation (3 of
4	2361	74.6	110000	12	BX322642_2	Continuation (3 of
c 5	2361	74.6	22001	6	AL772303	AL772303 Mouse DNA
c 6	1408.4	44.5	235283	12	AC120773	AC120773 Rattus no
7	1408.4	44.5	264245	12	AC096804	AC096804 Rattus no
c 8	1028	32.5	1197	6	AJ699423	AJ699423 Rattus no
9	861.8	27.2	1234	5	AJ621583	AJ621583 Pan trogl
10	857.8	27.1	1197	5	AJ697663	AJ697663 Pan trogl
11	757	23.9	1197	14	AJ868431	AJ868431 Bos tauru
12	559.8	17.7	1149	5	CQ736679	CQ736679 Sequence
c 13	470.4	14.9	151289	5	AL158164	AL158164 Human DNA
14	467.2	14.8	1197	11	AJ699424	AJ699424 Gallus ga
15	456.8	14.4	1520	11	CR352448	CR352448 Gallus ga
16	381	12.0	112099	12	AC143457	AC143457 Macaca mu
17	257.8	8.1	493	7	BV211170	BV211170 SIAT8F_5
18	256.6	8.1	1334	11	AJ704564	AJ704564 Gallus ga
19	248.2	7.8	1140	11	AJ715548	AJ715548 Tetraodon
20	247.6	7.8	1131	6	AJ699422	AJ699422 Rattus no
21	244	7.7	1131	5	AJ697662	AJ697662 Pan trogl
22	244	7.7	1131	5	CR457037	CR457037 Homo sapi
23	244	7.7	1878	2	CQ721379	CQ721379 Sequence
24	244	7.7	1953	5	BC108910	BC108910 Homo sapi
25	244	7.7	1953	5	BC108911	BC108911 Homo sapi
26	242.8	7.7	1881	6	BC034855	BC034855 Mus muscu
27	242.4	7.6	2594	5	AK056270	AK056270 Homo sapi
28	241.2	7.6	1854	6	X98014	X98014 M.musculus
29	239.2	7.6	1899	5	U91641	U91641 Human alpha
30	237.8	7.5	2135	5	AK094273	AK094273 Homo sapi
31	234.4	7.4	1125	14	AJ868432	AJ868432 Bos tauru
32	231.6	7.3	2094	5	AK130126	AK130126 Homo sapi
33	228.8	7.2	1032	11	AJ715546	AJ715546 Danio rer
34	228.2	7.2	1137	11	AJ871609	AJ871609 Oryzias l
35	224.4	7.1	1071	2	AX481522	AX481522 Sequence
36	224.4	7.1	1630	5	HUMCD3G	L43494 Human gangl
37	224.4	7.1	1704	5	HUMCD3S	D26360 Homo sapien
38	224.4	7.1	1755	2	CQ721891	CQ721891 Sequence
39	224.4	7.1	2117	5	AR360634	AR360634 Sequence
40	224.4	7.1	2117	5	HSG3S	X77922 H.sapiens G
41	222.8	7.0	1650	5	HUMBSIATR	L32867 Homo sapien
42	221.2	7.0	1071	5	AJ697658	AJ697658 Pan trogl
43	220.4	7.0	1116	11	AJ705092	AJ705092 Takifugu
44	220.4	7.0	1140	11	AJ715547	AJ715547 Takifugu
45	211.6	6.7	2381	6	BC024821	BC024821 Mus muscu

ALIGNMENTS

```
RESULT 1
AB059554
LOCUS       AB059554                3166 bp    mRNA    linear   ROD 27-APR-2005
DEFINITION  Mus musculus ST8Sia VI mRNA for alpha 2,8-sialyltransferase,
            complete cds.
ACCESSION   AB059554
VERSION     AB059554.1  GI:21668466
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Takashima,S., Ishida,H.K., Inazu,T., Ando,T., Ishida,H., Kiso,M.,
            Tsuji,S. and Tsujimoto,M.
  TITLE     Molecular cloning and expression of a sixth type of alpha
            2,8-sialyltransferase (ST8Sia VI) that sialylates O-glycans
  JOURNAL   J. Biol. Chem. 277 (27), 24030-24038 (2002)
  PUBMED    11980897
REFERENCE   2  (bases 1 to 3166)
  AUTHORS   Takashima,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-APR-2001) Shou Takashima, RIKEN, Cellular
            Biochemistry Laboratory; 2-1 Hirosawa, Wako, Saitama, 351-0198,
            Japan (E-mail:staka@riken.jp, Tel:81-48-462-1111(ex.3424),
            Fax:81-48-462-4670)
FEATURES             Location/Qualifiers
     source          1..3166
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /country="Japan"
     gene            1..3166
                     /gene="ST8Sia VI"
     CDS             77..1273
                     /gene="ST8Sia VI"
                     /codon_start=1
                     /product="alpha 2,8-sialyltransferase"
                     /protein_id="BAC01265.1"
                     /db_xref="GI:21668467"
```

```
        /translation="MRSGGTLFALIGSLMLLLLLRMLWCPADAPARSRLLMEGSREDT
        SGTSAALKTLWSPTTPVPRTRNSTYLDEKTYQITEKCKDLQYSLNSLSNKTRRYSEDD
        YLQTITNIQRCPWNRQAEEYDNFRAKLASCCDAIQDFVVSQNNTPVGTNMSYEVESKK
        HIPIRENIFRMFPVSQPFVDYPYNOCAVVGNGGILNKSLCGAEIDKSDFVFRCNLPPI
        TGSASKDVGSKTNLVTVNPSIITLKYQNLKEKKAQPLEDISTYGDAFLLLPAFSYRAN
        TGISFKVYQTLKESKMRQKYLFFHPRYLRHLALFWRTKGVTAYRLSTGLMIASVAVEL
        CENVKLYGFWPFSKTIEDTPLSHHYYDNMLPKHGFHQMPKEYSQMLQLIHMRGILKLQF
        SKCETA"

ORIGIN

Query Match           100.0%;   Score 3166;   DB 6;   Length 3166;
Best Local Similarity 100.0%;   Pred. No. 0;
Matches 3166;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;
```

Left column (residues 1–900):

```
Qy     1 CGGAGCGGCGAGTCGGTGCCGCCCGGCTTCGCCCCGGCAGCTTTGGCGGCGAGG       60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 CGGAGCGGCGAGTCGGTGCCGCCCGGCTTCGCCCCGGCAGCTTTGGCGGCGAGG       60

Qy    61 ACGCCCGTGGCTCAGGATGAGATCGGGGGACACGCGTGTTCGCCCTCATAGGCAGCCTGAT  120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 ACGCCCGTGGCTCAGGATGAGATCGGGGGACACGCGTGTTCGCCCTCATAGGCAGCCTGAT  120

Qy   121 GCTGCTGTCCTCCTGGTATGCTGTCTGTGCCCAGCCGCCCTGCCCGCCTTCCAGGCT      180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GCTGCTGTCCTCCTGGTATGCTGTCTGTGCCCAGCCGCCCTGCCCGCCTTCCAGGCT      180

Qy   181 GTTGATGGAGGAAGCAGAGGACACCAGTGTACCTCAGTGCACTGAAGACACTCTG        240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GTTGATGGAGGAAGCAGAGGACACCAGTGTACCTCAGTGCACTGAAGACACTCTG        240

Qy   241 GAGCCCGACACACCCCGTACCACGCACCAGGACACAGCACATATCTGATGATGAAGAAGACAAC 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GAGCCCGACACACCCCGTACCACGCACCAGGACACAGCACATATCTGATGATGAAGAAGACAAC 300

Qy   301 CCAAATAACAGAGAAATGCAAAGATCTGCAATATAGCTTGAACTCTTTATCTCAACAAAAC  360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CCAAATAACAGAGAAATGCAAAGATCTGCAATATAGCTTGAACTCTTTATCTCAACAAAAC  360

Qy   361 GAGACGTACTCTGAGGATGACTACCTCCAGACATCACAAACATCACAGATGCCCATG      420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 GAGACGTACTCTGAGGATGACTACCTCCAGACATCACAAACATCACAGATGCCCATG      420

Qy   421 GAACCCGGCAAGCAAGAGAAGAATATGACAATTTGAGAGCAAATGGCTTCCTGGTTGCGATGC 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GAACCCGGCAAGCAAGAGAAGAATATGACAATTTGAGAGCAAATGGCTTCCTGGTTGCGATGC 480

Qy   481 CATTCAAGACTTCGTGGTTCCAGAACAACTCCAGTGGGGACTAACATGAGCTACGA      540
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 CATTCAAGACTTCGTGGTTCCAGAACAACTCCAGTGGGGACTAACATGAGCTACGA      540

Qy   541 GGTGGAAAGCAAGAAACACATCCCATTTGAGAGAACATTTTCCACATGTTCCAGTGTC     600
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 GGTGGAAAGCAAGAAACACATCCCATTTGAGAGAACATTTTCCACATGTTCCAGTGTC     600

Qy   601 GCAGCTTTGTGGACTATCCCTATAACCAGTGTGCGCAGTGGTGGTAATGTGCCTG       660
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 GCAGCTTTGTGGACTATCCCTATAACCAGTGTGCGCAGTGGTGGTAATGTGCCTG       660

Qy   661 CAACAAGTCTCTCTGCGGAGCGCTAGTAAAGATGTGGAAGCAAAACAAATCTTGTGACTGT   720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 CAACAAGTCTCTCTGCGGAGCGCTAGTAAAGATGTGGAAGCAAAACAAATCTTGTGACTGT   720

Qy   721 CCCCCAATCACAGGAGCGCTAGTAAAGATGTGGAAGCAAAACAAATCTTGTGACTGT      780
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 CCCCCAATCACAGGAGCGCTAGTAAAGATGTGGAAGCAAAACAAATCTTGTGACTGT      780

Qy   781 CAATCCAGCATTATAACCCTGAAGTACCAGAATTTGAAGGAGAAGAAGCACAGTTTTT     840
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 CAATCCAGCATTATAACCCTGAAGTACCAGAATTTGAAGGAGAAGAAGCACAGTTTTT     840

Qy   841 GGAGGACATCTCCACCTATGGAGATGCATTCCTCCTGCCAGCATTTCCTATCGGGC      900
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 GGAGGACATCTCCACCTATGGAGATGCATTCCTCCTGCCAGCATTTCCTATCGGGC      900
```

Right column (residues 901–1980):

```
Qy   901 CAACACAGGCATCTCTTTTAAAACCTACCAAACACTCAAAGAGTCAAAAATGAGGCAAAA    960
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 CAACACAGGCATCTCTTTTAAAACCTACCAAACACTCAAAGAGTCAAAAATGAGGCAAAA    960

Qy   961 GGTTCTCTTCTTCCATCCCAGGAACCTGAGAACACACCTCGCTCTTTTCTGGAGAACTAAAGG 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 GGTTCTCTTCTTCCATCCCAGGAACCTGAGAACACACCTCGCTCTTTTCTGGAGAACTAAAGG 1020

Qy  1021 GGTGACTGACTGCATACCGCTGTGTCCCAGGCTTGATGATTGCAAGTGTCGCTGTGGAACTGTG 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021 GGTGACTGACTGCATACCGCTGTGTCCCAGGCTTGATGATTGCAAGTGTCGCTGTGGAACTGTG 1080

Qy  1081 TGAAAACGTGAAGCTCACGATCTGGCCTTTCTCTAAGACTATCGAAGACACCCCACT    1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081 TGAAAACGTGAAGCTCACGATCTGGCCTTTCTCTAAGACTATCGAAGACACCCCACT    1140

Qy  1141 CAGTCACCACTACTATGATAACTGTTACCTAAGCATGGTTCCACCAGATGCCTAAAGA   1200
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141 CAGTCACCACTACTATGATAACTGTTACCTAAGCATGGTTCCACCAGATGCCTAAAGA   1200

Qy  1201 ATACAGCAAATGCTCCAGCTCAATGAGAGAATTCTCAAACTGCAATTCAGCAAATG    1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201 ATACAGCAAATGCTCCAGCTCAATGAGAGAATTCTCAAACTGCAATTCAGCAAATG    1260

Qy  1261 TGAAACGGCTTAACGTTCTTAGAGGAGAATAATTTCAGGAGGTGGAGTGGATGTGTCA   1320
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261 TGAAACGGCTTAACGTTCTTAGAGGAGAATAATTTCAGGAGGTGGAGTGGATGTGTCA   1320

Qy  1321 CAGCATCTTCCAAAAAGCCAATACAGAGAAAAGCATGAATGAAAGGCACAGAGGCGCT   1380
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321 CAGCATCTTCCAAAAAGCCAATACAGAGAAAAGCATGAATGAAAGGCACAGAGGCGCT   1380

Qy  1381 CTCCCACTTGTCTGAGACCAAGTCTCTTCTCTGAGAATAGAGACCAAAACATCAGACTTGGAT 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381 CTCCCACTTGTCTGAGACCAAGTCTCTTCTCTGAGAATAGAGACCAAAACATCAGACTTGGAT 1440

Qy  1441 CTCATTCTCACCTTCAACGTTCAATCATCATAGAATTTCAGATGAGCCAGTTTCCAGAGTGA 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1441 CTCATTCTCACCTTCAACGTTCAATCATCATAGAATTTCAGATGAGCCAGTTTCCAGAGTGA 1500

Qy  1501 AAGTAAAATGAGATAATTTTCAGATGAGCCAGTTTCCACCTTTATCAGAGTGGTAATGAA   1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1501 AAGTAAAATGAGATAATTTTCAGATGAGCCAGTTTCCACCTTTATCAGAGTGGTAATGAA   1560

Qy  1561 GAATGCTCCTTGTTCCTATCCGATAGCCATTCCCACCTTTATCAGAGTGGTAATGAA    1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1561 GAATGCTCCTTGTTCCTATCCGATAGCCATTCCCACCTTTATCAGAGTGGTAATGAA    1620

Qy  1621 ACTGTGCAATTGTGCCAAAGACCTTTCTGAAGAGAATGTCTGAATCATGCGCCGAGTTT   1680
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1621 ACTGTGCAATTGTGCCAAAGACCTTTCTGAAGAGAATGTCTGAATCATGCGCCGAGTTT   1680

Qy  1681 TTACACACAGCTCTTCCTTTATAATAAATCCTTCCATTCTCCCTCCTCCTTAGTGCCTTCTA    1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1681 TTACACACAGCTCTTCCTTTATAATAAATCCTTCCATTCTCCCTCCTCCTTAGTGCCTTCTA    1740

Qy  1741 GAAACAAAATACCCTTGATGATCAGGAAGAAAGTCTTTTTTACTTAGCAATGTGCCTG   1800
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1741 GAAACAAAATACCCTTGATGATCAGGAAGAAAGTCTTTTTTACTTAGCAATGTGCCTG   1800

Qy  1801 CTTCTGATTCAGTTCGCTTGTGCATTAAGCTGGGTTTGGGGTTTGGGATTTGGGGC    1860
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1801 CTTCTGATTCAGTTCGCTTGTGCATTAAGCTGGGTTTGGGGTTTGGGATTTGGGGC    1860

Qy  1861 GTTTCTTCACTTCTTTTGTCTATTTCCTTACCTTTATCAGTTTGTATTCGAGCTTCC    1920
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1861 GTTTCTTCACTTCTTTTGTCTATTTCCTTACCTTTATCAGTTTGTATTCGAGCTTCC    1920

Qy  1921 TGCTTTGGGATTCTGCAATTCTCCTCCCCACTGACAGGATCAACTCAATGACATAAAGTA   1980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1921 TGCTTTGGGATTCTGCAATTCTCCTCCCCACTGACAGGATCAACTCAATGACATAAAGTA   1980
```

Qy 1981 GTTCAAACATCCATTGCTTCTCACACGTTTATCCATACAAGTTACTCATCTGATTTTATT 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1981 GTTCAAACATCCATTGCTTCTCACACGTTTATCCATACAAGTTACTCATCTGATTTTATT 2040

Qy 2041 TAAAATAGTGAACATCTCACTTGATATCAGACCCGAGACCATCCTCCATTGGAGAATATG 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2041 TAAAATAGTGAACATCACTTGATATCAGACCCGAGACCATCCTCCATTGGAGAATATG 2100

Qy 2101 AAGATATTGTCACTGGCAGAGAAAAGCAGGTGTGTGCCATTAATTGATAAGATACCACAGC 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2101 AAGATATTGTCACTGGCAGAGAAAAGCAGGTGTGTGCCATTAATTGATAAGATACCACAGC 2160

Qy 2161 ATCATCATGCCAGTTATGAACACAGTCCTGAAAGGATCATAGACAGGGTGTTAAATCT 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2161 ATCATCATGCCAGTTATGAACACAGTCCTGAAAGGATCATAGACAGGGTGTTAAATCT 2220

Qy 2221 GATCCCAGTAGAATAAACTTCAGTGTACCTATTTCAGGGAAGAGTTAAATTTCACAATTAA 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2221 GATCCCAGTAGAATAAACTTCAGTGTACCTATTTCAGGGAAGAGTTAAATTTCACAATTAA 2280

Qy 2281 AACTAGTAAATGAACCAATTCTTAGGCACATTAAGTGGATTCTGAGTAAAAGAAAGGGAA 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2281 AACTAGTAAATGAACCAATTCTTAGGCACATTAAGTGGATTCTGAGTAAAAGAAAGGGAA 2340

Qy 2341 CAGCAGGAGAAAGCTGTGCGCTGGTTCTGATTACCCAAATGAGCATGCTGGAAGGAGGT 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2341 CAGCAGGAGAAAGCTGTGCGCTGGTTCTGATTACCCAAATGAGCATGCTGGAAGGAGGT 2400

Qy 2401 TGTGAGGCTACGCTAAAACCTCGCGTAGGGAGAGTACAGTGCATGAGTGTGCGCGCT 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2401 TGTGAGGCTACGCTAAAACCTCGCGTAGGGAGAGTACAGTGCATGAGTGTGCGCGCT 2460

Qy 2461 TTTGTCCACACTCGTGAAGGGTGAGTAATTCAGAGCCAATCACATCACAAGGATGACAC 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2461 TTTGTCCACACTCGTGAAGGGTGAGTAATTCAGAGCCAATCACATCACAAGGATGACAC 2520

Qy 2521 ACCTAACTCATCACTTCAGGGGAGATGAAGTGCTTTCATGAGAAATTACACTCATAGCT 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2521 ACCTAACTCATCACTTCAGGGGAGATGAAGTGCTTTCATGAGAAATTACACTCATAAGCT 2580

Qy 2581 AAGCATCAGTTTGAGTAAAATTTGAGTAGATGTTAAATATGAACATTTTATACCTCTTA 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2581 AAGCATCAGTTTGAGTAAAATTTGAGTAGATGTTAAATATGAACATTTTATACCTCTTA 2640

Qy 2641 CTAAATGTCCCACCGACACCTTTAATGTAAGCACATTTATTAAGTTACTTGACATT 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2641 CTAATGTCCCACCGACACCTTTAATGTAAGCACATTTATTAAGTTACTTGACATT 2700

Qy 2701 AAATGCTTATGTCTGTATATTCTGTTCATCCATCGATTTTCCCAAAAGTAAGAGCATAG 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2701 AAATGCTTATGTCTGTATATTCTGTTCATCCATCGATTTTCCCAAAAGTAAGAGCATAG 2760

Qy 2761 GAGATGAGGCCTACATGCCAAGAAAACTATAAATTTACTCTTAATTCTTACTTGAGCC 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2761 GAGATGAGGCCTACATGCCAAGAAAACTATAAATTTACTCTTAATTCTTACTTGAGCC 2820

Qy 2821 AGCTTGTTGTTTATCAAGTGCTTTTTGAAGAGACAGCACCCTGTGAATTCTTCATTCTG 2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2821 AGCTTGTTGTTTATCAAGTGCTTTTTGAAGAGACAGCACCCTGTGAATTCTTCATTCTG 2880

Qy 2881 ATACAGTGTCACCCTTGTATTTAACATTTGTAATGTTGTTGTTTCAAGTTTACATCTCTTTCAT 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2881 ATACAGTGTCACCCTTGTATTTAACATTTGTAATGTTGTTTCAAGTTTACATCTCTTTCAT 2940

Qy 2941 TCTTTTATAGCAAATCAAACGTATTAGCTTCAGAAATTATCAGAAGTTCATATATAAAT 3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2941 TCTTTTATAGCAAATCAAACGTATTAGCTTCAGAAATTATCAGAAGTTCATATATAAAT 3000

Qy 3001 ATTTTGCAAAGGGTAAAAGGCTTTTTTGTTAAATAAAATAAATTATTTATTTTCTTCTG 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3001 ATTTTGCAAAGGGTAAAAGGCTAAAAGGCTTTTTTGTTAAATAAAATAAATTATTTATTTTCTTCTG 3060

Qy 3061 ATGAATAGAGGCTCTTTATGCTGCTCGTAATGAACCTAATTAGCTTTAAATTATCTCCT 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 3061 ATGAATAGAGGCTCTTTATGC CTGCTCAATGAACCTAATTAGCTTTAAATTATCTCCT 3120

Qy 3121 AGCAACATTGGTCACGTTCAA CATGCTATTAGCAAAAAAAAAAA 3166
          ||||||||||||||||||||| ||||||||||||
Db 3121 AGCAACATTGGTCACGTTCAA CATGCTATTAGCAAAAAAAAAAA 3166

RESULT 2
BX322642 0/c
WPCOMMENT
       Sequence split into 5 fragments   LOCUS BX322642 Accession BX322642
          Fragment Name   , Begin   End
          BX322642_0   1   110000
          BX322642_1   100001   210000
          BX322642_2   200001   310000
          BX322642_3   300001   410000
          BX322642_4   400001   440302

LOCUS   BX322642   440302 bp   DNA   linear   HTG 03-FEB-2006
DEFINITION Mus musculus chromosome 2 clone RP24-36309, WORKING DRAFT SEQUENCE,
            8 unordered pieces.
ACCESSION   BX322642
VERSION   BX322642.14 GI:86476683
KEYWORDS   HTG; HTGS PHASE1; HTGS ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 440802)
AUTHORS   Lloyd,D.
TITLE   Direct Submission
JOURNAL   Submitted (02-FEB-2006) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
            On Feb 4, 2006 this sequence version replaced gi:76666521.
COMMENT   -- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            -- Project Information
            Center project name: zKp57E9
            -- Summary Statistics
            Assembly program: XGAP4;
            Chemistry: Dye-terminator; version 4.5
            Consensus quality: 272872 100% of reads
            Consensus quality: 272872 bases at least Q40
            Consensus quality: 274462 bases at least Q30
            Consensus quality: 275395 bases at least Q20
            Insert size: 440102; sum-of-contigs
            Insert size: 172000; 10.8 error; agarose-fp
            Quality coverage: 6.40x in Q20 bases; sum-of-contigs Quality
            coverage: 16.59x in Q20 bases; agarose-fp
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *   1   163997: contig of 163997 bp in length
            *   163998   164097: gap of 100 bp
            *   164098   166408: contig of 2311 bp in length
            *   166409   166509: gap of 101 bp
            *   166510   170049: contig of 3540 bp in length
            *   170050   170150: gap of 101 bp
            *   170151   214018: contig of 43868 bp in length
            *   214019   214118: gap of 100 bp
            *   214119   232782: contig of 18664 bp in length
            *   232783   232884: gap of 102 bp
            *   232885   258463: contig of 25579 bp in length

```
FEATURES             Location/Qualifiers
     *      258464    258563: gap of 100 bp
     *      258564    276703: contig of 18140 bp in length
     *      276704    276803: gap of 100 bp
     *      276804    440802: contig of 163999 bp in length.
     source        1..440802
                    /organism="Mus musculus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10090"
                    /chromosome="2"
                    /clone="RP24-36309"
                    /clone_lib="RPCI-24"
     misc_feature  1..163997
                    /note="assembly_fragment:00420
     misc_feature  164098..166408
                    clone_end:SP6
                    vector_side:right
                    clone_end:T7
                    vector_side:left"
     misc_feature  166509..170049
                    /note="assembly_fragment:00054"
     misc_feature  170150..214018
                    /note="assembly_fragment:04636"
     misc_feature  214119..232782
                    /note="assembly_fragment:04637"
     misc_feature  232883..258463
                    /note="assembly_fragment:04642"
     misc_feature  258564..276703
                    /note="assembly_fragment:04643"
     misc_feature  276804..440802
                    /note="assembly_fragment:04645"
                    /note="assembly_fragment:04648"

ORIGIN
Query Match         74.6%;     Score 2361;   DB 12;   Length 110000;
Best Local Similarity   99.6%;   Pred. No. 0;
Matches 2367;  Conservative  0;  Mismatches  10;  Indels  0;  Gaps  0;
```

(Sequence alignment between Query (Qy) and Database (Db) follows)

```
Qy   2410  AGCTAAAACCTGCGTAGGGAGAGAGTACAGTGCATGAGTGTGGCGGCTTTTGTCCAC  2469
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  20083  AGCTAAAACCCTGCGTAGGGAGAGAGTACAGTGCATGAGTGTGGCGGCTTTTGTCCAC  20024

Qy   2470  ACTCGTGAAGGGTGAGTAATTCAGAGCCAATCACATCACAAGGATGACACACCTAACTC  2529
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  20023  ACTCGTGAAGGGTGAGTAATTCAGAGCCAATCACATCACAAGGATGACACACCTAACTC  19964

Qy   2530  ATCACTTCAGGGGAGGAGATGAATGCTTTCATGAGAGAATTACACTCATAAGCATCAG  2589
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19963  ATCACTTCAGGGGAGGAGATGAATGCTTTCATGAGAGAATTACACTCATAAGCATCAG  19904

Qy   2590  TTTTGAGTAAAATTGAGTAGAGTGTAAATATGAACACATTTATACCTCTTACTAAATGTCC  2649
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19903  TTTTGAGTAAAATTGAGTAGAGTGTAAATATGAACACATTTATACCTCTTACTAAATGTCC  19844

Qy   2650  CACCGACACCTTTAATGTAAGCACACATTTATTTATTAAGTACTTGACACATTAAATGCTTA  2709
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19843  CACCGACACCTTTAATGTAAGCACACATTTATTTATTAAGTACTTGACACATTAAATGCTTA  19784

Qy   2710  TGTCTGTATATTCTGTTCATCATCGATTTCCCAAAAAAGTAAGAGCATAGGAGATGAGG  2769
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19783  TGTCTGTATATTCTGTTCATCATCGATTTCCCAAAAAAGTAAGAGCATAGGAGATGAGG  19724

Qy   2770  CCTACATGCCAAGAAAACTATAAATTTACTCTTAATTCTTAATTCTTGAGCCAGCTTGTG  2829
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19723  CCTACATGCCAAGAAAACTATAAATTTACTCTTAATTCTTAATTCTTGAGCCAGCTTGTG  19664

Qy   2830  TTTATCAAGTGCTTTTTGAAGAGAGCAGCACCCTGTGAATTCTTCATTCTGATACAGTGT  2889
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19663  TTTATCAAGTGCTTTTTGAAGAGAGCAGCACCCTGTGAATTCTTCATTCTGATACAGTGT  19604

Qy   2890  CACCTTGTATTTAAACATTGTAATGTGTTTCAAGTTTACATCTCTTTCATTCTTTTATA  2949
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19603  CACCTTGTATTTAAACATTGTAATGTGTTTCAAGTTTACATCTCTTTCATTCTTTTATA  19544

Qy   2950  GCAAATCAAAGTATTAGCTTCAGAAATTTATCAGAAGTTCATATATAAATATTTTGCAA  3009
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19543  GCAAATCAAAGTATTAGCTTCAGAAATTTATCAGAAGTTCATATATAAATATTTTGCAA  19484

Qy   3010  AGGGTAAAAGGCTTTTTGTTTGTTGTTAAATAAAATAAAATTTATTATTTCTTCTGATGAATAGA  3069
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19483  AGGGTAAAAGGCTTTTTGTTTGTTGTTAAATAAAATAAAATTTATTATTTCTTCTGATGAATAGA  19424

Qy   3070  GGCTCTTTTATGCTGCTAATGAACCTAATTAGCTTTAAATATCTCCTAGCAACATT  3129
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19423  GGCTCTTTTATGCTGCTAATGAACCTAATTAGCTTTAAATATCTCCTAGCAACATT  19364

Qy   3130  GGTCACGTTCAATCATGCTATTAGCACAAAAAAAAAAAA  3166
           |||||||||||||||||||||||||||||||||| ||||
Db  19363  GGTCACGTTCAATCATGCTATTAGCACAAAAAAAAGAAAA  19327
```

RESULT 3
BX322642_2
WPCOMMENT
Sequence split into 5 fragments   LOCUS BX322642 Accession BX322642

Fragment Name	Begin	End
BX322642_0	1	110000
BX322642_1	100001	210000
BX322642_2	200001	310000
BX322642_3	300001	410000
BX322642_4	400001	440802

Continuation (3 of 5) of BX322642 from base 200001 (BX322642 Mus musculus chromosome 2 c

Query Match   74.6%;   Score 2361; DB 12; Length 110000;
Best Local Similarity 99.6%;   Pred. No. 0;
Matches 2367;   Conservative 0;   Mismatches 10;   Indels 0;   Gaps 0;

```
Qy    790  CATTATAACCCTGAAGTACCAGAATTTGAAGGAGAAGAAAGCACAGTTTTTGGAGGACAT  849
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36760  CTTTTAATTCGGCAGTGACAGTTACCAGAATTTGAAGGAGAAGAAAGCACAGTTTTTGGAGGACAT  36819
```

```
Qy    850  CTCCACCTATGGAGAATGCATTC    909
           |||||||||||||||||||||||
Db  36820  CCCCTCCTGCCAGCATTTCCTATCGGGCCAACACAGG  36879

Qy    910  CATCTCTTTAAAGTCTACCAA    969
           |||||||||||||||||||||
Db  36880  CACTCTTTTAAAGTCTACCAA  36939

Qy    970  CTTCCATCCCAGTACTTGAGA   1029
           |||||||||||||||||||||
Db  36940  CTTCCATCCCAGTACTTGAGA  36999

Qy   1030  ATACCGCTTGTCCACAGGCTTG   1089
           ||||||||||||||||||||||
Db  37000  ATACCGCTTGTCCACAGGCTTG  37059

Qy   1090  GAAGCTCTACGGATTCTGGCCT   1149
           |||||||||||||||||||||
Db  37060  GAAGCTCTACGGATTCTGGCCT  37119

Qy   1150  CTACTATGATAACATGTTACCT   1209
           |||||||||||||||||||||
Db  37120  CTACTATGATAACATGTTACCT  37179

Qy   1210  AATGCTCCAGCTCCATATGAGA   1269
           |||||||||||||||||||||
Db  37180  AATGCTCCAGCTCCATATGAGA  37239

Qy   1270  TTAACGTTTCTTAGAAGGAGAA   1329
           |||||||||||||||||||||
Db  37240  TTAACGTTTCTTAGAAGGAGAA  37299

Qy   1330  CAAAAAGCAATAGAAGAAGGC   1389
           |||||||||||||||||||||
Db  37300  CAAAAAGCAATAGAAGAAGGC  37359

Qy   1390  GTCTAGACCAAAGCCACCCGCC   1449
           |||||||||||||||||||||
Db  37360  GTCTAGACCAAAGCCACCCGCC  37419

Qy   1450  ACCTTCAACGTTCTTTCTGA   1509
           ||||||||||||||||||||
Db  37420  ACCTTCAACGTTCTTTCTGA  37479

Qy   1510  GAGATAATTTTCAAATCATCA   1569
           |||||||||||||||||||||
Db  37480  GAGATAATTTTCAAATCATCA  37539

Qy   1570  CTTGTTCCTATCCATGATAGCC   1629
           |||||||||||||||||||||
Db  37540  CTTGTTCCTATCCATGATAGCC  37599

Qy   1630  TTGTGCCAAGACCCTTTCTGA   1689
           |||||||||||||||||||||
Db  37600  TTGTGCCAAGACCCTTTCTGA  37659

Qy   1690  GCTCTTCCTTTATAAATAAATC   1749
           |||||||||||||||||||||
Db  37660  GCTCTTCCTTTATAAATAAATC  37719

Qy   1750  TACCCTTGATGATTCAGGAAGA   1809
           |||||||||||||||||||||
Db  37720  TACCCTTGATGATTCAGGAAGA  37779

Qy   1810  CAGTTCGCTTGTGACATTAAGC   1869
           |||||||||||||||||||||
Db  37780  CAGTTCGCTTGTGACATTAAGC  37839

Qy   1870  CTTCTTTTGTCTATATTTTCCT   1929
           |||||||||||||||||||||
Db  37840  CTTCTTTTGTCTATATTTTCCT  37899

Qy   1930  ATTCTGCAATTCTCTCTCCCAC   1989
```

```
Db  38980  AGGTAAAAGGCTTTTGTTA ... ATAAATAAAATTATTATTTCTTCGATGAATAGA  39039

Qy   3070  GGCTCTTTATGCTGCTGCTAA ... GAACCTAATTAGCTTAAATTATCTCCTAGCAACATT  3129
           |||||||||||||||||||||       |||||||||||||||||||||||||||||||||
Db  39040  GGCTCTTTATGCTGCTGCTAA ... GAACCTAATTAGCTTAAATTATCTCCTAGCAACATT  39099

Qy   3130  GGTCACGTTCAATCATGCTAT ... AGCAAAAAAAAAA  3166
           |||||||||||||||||||||       |||| ||||
Db  39100  GGTCACGTTCAATCATGCTAT ... GCAAAAAAAGAAAA  39136


RESULT 4
BX322642_2/c
WPCOMMENT
LOCUS BX322642  Accession BX322642

Sequence split into 5 fragments
   Fragment Name      Begin        End
   BX322642_0             1      110000
   BX322642_1        100001      210000
   BX322642_2        200001      310000
   BX322642_3        300001      410000
   BX322642_4        400001      440302
Continuation (3 of 5) of BX322642 from base 200001 (BX322642 Mus musculus chromosome 2 ...

Query Match  74.6%;   Score 2361;  DB 12;   Length 110000;
Best Local Similarity 99.6%;  Pred. No. 0;
Matches 2367;  Conservative 0;  Mismatches 10;  Indels 0;  Gaps 0;

Qy    790  CATTATAACCTGAAGTACCAG ... ATTTGAAGGAGAAGAAAGCACAGTTTTTGGAGGACAT   849
           | | | |||||||||||||||       |||||||||||||||||||||||||||||||||||
Db  98506  CTTTTAATTCGGCAGTACCAG ... TTTTGAAGGAGAAGAAAGCACAGTTTTTGGAGGACAT  98447

Qy    850  CTCCACCTATGGAGATGCATTCC ... TCCTCCTGCCAACACAGG   909
           |||||||||||||||||||||||       |||||||||||||||||
Db  98446  CTCCACCTATGGAGATGCATTCC ... TCCTCCTGCCAACACAGG  98387

Qy    910  CATCTCTTTAAAGTCTACCAA ... CACTCAAAGAGTCAAAAGTTCTCTT   969
           |||||||||||||||||||||       |||||||||||||||||||||||||
Db  98386  CATCTCTTTAAAGTCTACCAA ... CACTCAAAGAGTCAAAAGTTCTCTT  98327

Qy    970  CTTCCATCCCAGTACCTGAGA ... ACCTCGCTCTTTTCTGGAGAACTAAAGGGGTGACTGC  1029
           |||||||||||||||||||||       |||||||||||||||||||||||||||||||||||
Db  98326  CTTCCATCCCAGTACCTGAGA ... ACCTCGCTCTTTTCTGGAGAACTAAAGGGGTGACTGC  98267

Qy   1030  ATACCGCTTGTCCACAGGCTTGA ... TGATTGCAAGTGTCGCTGTGGAACTGTGTGAAAAGT  1089
           |||||||||||||||||||||||       |||||||||||||||||||||||||||||||||
Db  98266  ATACCGCTTGTCCACAGGCTTGA ... TGATTGCAAGTGTCGCTGTGGAACTGTGTGAAAACGT  98207

Qy   1090  GAAGCTTCCAGGATTCGGGCCTT ... TCCTAAGACTATCGAAGACACCCCACTCAGTCACCA  1149
           |||||||||||||||||||||||       |||||||||||||||||||||||||||||||||
Db  98206  GAAGCTTCCAGGATTCGGGCCTT ... TCCTAAGACTATCGAAGACACCCCACTCAGTCACCA  98147

Qy   1150  CTACTATGATAACATGTTACCTA ... AGCATGGTTTCCACCAGATGCCTAAGAATACAGCCA  1209
           |||||||||||||||||||||||       |||||||||||||||||||||||||||||||||
Db  98146  CTACTATGATAACATGTTACCTA ... AGCATGGTTTCCACCAGATGCCTAAGAATACAGCCA  98087

Qy   1210  AATGCTTCCAGCTCCATATGAGA ... GAATCCTCAAACTGCAATTCAGCAAATGTGAAACGGC  1269
           |||||||||||||||||||||||       ||||||||||||||||||||||||||||||||||
Db  98086  AATGCTTCCAGCTCCATATGAGA ... GAATCCTCAAACTGCAATTCAGCAAATGTGAAACGGC  98027

Qy   1270  TTAACGTTCTTCTTAGAGGAGAA ... CAATTTCAGGAGGTGGAGTGGATGTGTCACAGCATCTC  1329
           |||||||||||||||||||||||       ||||||||||||||||||||||||||||||||||||
Db  98026  TTAACGTTCTTCTTAGAGGAGAA ... CAATTTCAGGAGGTGGAGTGGATGTGTCACAGCATCTC  97967

Qy   1330  CAAAAAGCCAATAGAAGAGAGCG ... CAGAGAAAGCAATGAATTACAAGAGGCGCTCTCCCACTT  1389
           |||||||||||||||||||||||       |||||||||||||||||||||||||||||||||||||
Db  97966  CAAAAAGCCAATAGAAGAGAGCG ... CAGAGAAAGCAATGAATTACAAGAGGCGCTCTCCCACTT  97907

Qy   1390  GTCTAGACCAAAGCCACCCGCC ... CCACTCACTTTGCAGCCTCCACGAGTCACTCATTCTC  1449
           |||||||||||||||||||||||       |||||||||||||||||||||||||||||||||||
Db  97906  GTCTAGACCAAAGCCACCCGCC ... CCACTCACTTTGCAGCCTCCACGAGTCACTCATTCTC  97847

Qy   1450  ACCTTCAACGTTCTTTCTCTGAC ... CATAGAGACCAAAAACATCAGACTTGGATAAGTAAAAT  1509
```

```
Db  37900  ... ATTCTGCAATTCTCTCCCACTGACAGATCAACTCAATGACATAAAGTAGTTCAAACA  37959
                                 ||||||||||||||||||||||||||||||||||||||||
Qy   1990  TCCATTGCTTCTTCACATGTTTATCCATAAAGTACTCATCTGATTTATTTAAAATAGT  2049
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  37960  TCCATTGCTTCTTCACATGTTTATCCATAAAGTACTCATCTGATTTATTTAAAATAGT  38019

Qy   2050  GAACATCTACTTGATATCAGAACCGAGGACCATCTCCTCCATTGGAGAATATGAAGATATTG  2109
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  38020  GAACATCTACTTGATATCAGAACCGAGGACCATCTCCTCCATTGGAGAATATGAAGATATTG  38079

Qy   2110  TCACTGGCAGAAAAGCAGGTGTGTGCCATTAATTGATAAGATACCACAAGCATCATCATG  2169
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  38080  TCACTGGCAGAAAAGCAGGTGTGTGCCATTAATTGATAAGATACCACAAGCATCATCATG  38139

Qy   2170  CCAGTTATGAACACAGTGCTGAAAGGATCATAGACAGGGTTGGTTAAATCTGATCCCAGT  2229
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  38140  CCAGTTATGAACACAGTGCTGAAAGGATCATAGACAGGGTTGGTTAAATCTGATCCCAGT  38199

Qy   2230  AGAATAAAACTTCAGTGTACCTATTCAGGGAAGAGTTAATTCACAATTAAAACTAGTAA  2289
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  38200  AGAATAAAACTTCAGTGTACCTATTCAGGGAAGAGTTAATTCACAATTAAAACTAGTAA  38259

Qy   2290  ATGAACCAATTCTTAGGCACATTAAGTGGATTCTGAGTGTAAAGAAAAGGGAACAGCAGGAG  2349
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  38260  ATGAACCAATTCTTAGGCACATTAAGTGGATTCTGAGTGTAAAGAAAAGGGAACAGCAGGAG  38319

Qy   2350  AAAGCTGTTCGCTTGGTTCGATTACCCAAATGAGCATGCTGAAGGAGGTTGTGAGGCT  2409
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  38320  AAAGCTGTTCGCTTGGTTCGATTACCCAAATGAGCATGCTGAAGGAGGTTGTGAGGCT  38379

Qy   2410  ACGCTAAAACCTCTGCGTAGGGAGGAGTACAGTGCATGGATGAGTGTGGCGGCTTTTGTCCAC  2469
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  38380  ACGCTAAAACCTCTGCGTAGGGAGGAGTACAGTGCATGGATGAGTGTGGCGGCTTTTGTCCAC  38439

Qy   2470  ACTCGTGAGGGTGAGTAATTCAGAGCCAATCACAAGGATGACACACCTAACTC  2529
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  38440  ACTCGTGAGGGTGAGTAATTCAGAGCCAATCACAAGGATGACACACCTAACTC  38499

Qy   2530  ATCACTTCAGGGAGAGATGAATGCTTTCATGAGAAATTACACTCATAAGCATCAG  2589
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  38500  ATCACTTCAGGGAGAGATGAATGCTTTCATGAGAAATTACACTCATAAGCATCAG  38559

Qy   2590  TTTTGAGTAAAATTGAGTAGTAAGCACATTTATTATTTATACCTCTTACTAATGTCC  2649
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  38560  TTTTGAGTAAAATTGAGTAGTAAGCACATTTATTATTTATACCTCTTACTAATGTCC  38619

Qy   2650  CACCGGACACCTTTAATGTAAGCACATTTGTTCATCATCGATTTCATTCTTTATA...  2709
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  38620  CACCGGACACCTTTAATGTAAGCACATTTGTTCATCATCGATTTCATTCTTTATA...  38679

Qy   2710  TGTCTGTATATTCTGTTCATCCATCGATTTTCTGATTTTCTGTTTACTTGACATTGAAGAGCATGAGG  2769
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  38680  TGTCTGTATATTCTGTTCATCCATCGATTTTCTGATTTTCTGTTTACTTGACATTGAAGAGCATGAGG  38739

Qy   2770  CCTACATGCCAAGAAAACTATAAATTTACTCCTTAATTCTTAATTCTTACTTGAGCCAGCTGTGTG  2829
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  38740  CCTACATGCCAAGAAAACTATAAATTTACTCCTTAATTCTTAATTCTTACTTGAGCCAGCTGTGTG  38799

Qy   2830  TTTATCAAGTGCTTTTGAGAGAAACAGCAGCAGCACCCTGAATTTCATTCTGATACAGTGT  2889
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  38800  TTTATCAAGTGCTTTTGAGAGAAACAGCAGCAGCACCCTGAATTTCATTCTGATACAGTGT  38859

Qy   2890  CACCTTGTATTTAACATTTGTAATGTGTTCAAGTTTACATCTCTTTCATTCTTTTATA  2949
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  38860  CACCTTGTATTTAACATTTGTAATGTGTTCAAGTTTACATCTCTTTCATTCTTTTATA  38919

Qy   2950  GCAAATCAAACGTATTAGCTTCAGAAATTTATCAGAAGTTCATATAAATATTTTGCAA  3009
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  38920  GCAAATCAAACGTATTAGCTTCAGAAATTTATCAGAAGTTCATATAAATATTTTGCAA  38979

Qy   3010  AGGGTAAAAGGCTTTTGTTAAATAAAATAAAATTATTATTTCTTCGATGAATAGA  3069
```

```
Db 97846 ACCTTCAAGTCTCTTCCTGAGAATAGAGACCAAAACATCAGACTTGGATAAGTAAAAT 97787

Qy  1510 GAGATAATTTTCAAATCATCATCAGAATTTGATTTGAGCCAGCCGGTCTCTCAGAGATGCTTC 1569
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 97786 GAGATAATTTTCAAATCATCATCAGAATTTGATTTGAGCCAGCCGGTCTCTCAGAGATGCTTC 97727

Qy  1570 CTTGTTCCTATCCATGATGAGCCATTCCCACCTTTATCAGAGTTGGTAATGAAACTGTGCAA 1629
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 97726 CTTGTTCCTATCCATGATGAGCCATTCCCACCTTTATCAGAGTTGGTAATGAAACTGTGCAA 97667

Qy  1630 TTGTGTGCCAAAGACCCCTTTCTGAAGAGAATGTCTGAATCATCGCCGCCAGTTTTTACACACA 1689
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 97666 TTGTGTGCCAAAGACCCCTTTCTGAAGAGAATGTCTGAATCATCGCCGCCAGTTTTTACACACA 97607

Qy  1690 GCTCTTCCTTTATAAATAAAATCCTTCCCATTCCCCTCCTAGTAGTACAGAGAAAACAAAA 1749
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 97606 GCTCTTCCTTTATAAATAAAATCCTTCCCATTCCCCTCCTAGTAGTACAGAGAAAACAAAA 97547

Qy  1750 TACCCTTGATGATTCAGGAAGAAAAGTCTTTTTTACTTAGCAATGTGCCTGCTTCTGATT 1809
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 97546 TACCCTTGATGATTCAGGAAGAAAAGTCTTTTTTACTTAGCAATGTGCCTGCTTCTGATT 97487

Qy  1810 CAGTTCGCTTGTGACATTAAGCTGGGTTGGGGGTTTGGTTGGATTTGGTTGGGCGTTTCTTCA 1869
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 97486 CAGTTCGCTTGTGACATTAAGCTGGGTTGGGGGTTTGGTTGGATTTGGTTGGGCGTTTCTTCA 97427

Qy  1870 CTTCTTTGTCTATATTTTCCTTATCAGTTGTATTCGAGCTTCGAGCTTCGAGCTTTGGG 1929
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 97426 CTTCTTTGTCTATATTTTCCTTATCAGTTGTATTCGAGCTTCGAGCTTCGAGCTTTGGG 97367

Qy  1930 ATTCTGCAATTCTCTCTCCCACTGACAGGATCAACTCAACTAAAGTAGTTCAAACA 1989
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 97366 ATTCTGCAATTCTCTCTCCCACTGACAGGATCAACTCAACTAAAGTAGTTCAAACA 97307

Qy  1990 TCCAATGCTTCTCACATGTTTTATCCATAAAGTACTCATCTGATTTTATTTAAAATAGT 2049
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 97306 TCCAATGCTTCTCACATGTTTTATCCATAAAGTACTCATCTGATTTTATTTAAAATAGT 97247

Qy  2050 GAACATCTACTTGATATCAGACCCGAGACCCGAGGACCACCTCCATTAATTGATAAGATACCA 2109
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 97246 GAACATCTACTTGATATCAGACCCGAGACCCGAGGACCACCTCCATTAATTGATAAGATACCA 97187

Qy  2110 TCACTGGCAGAAAAGCAGAGGTGTGTGCCATTAATTGATAAGAT 2169
         |||||||||||||||||||||||||||||||||||||||||||
Db 97186 TCACTGGCAGAAAAGCAGAGGTGTGTGCCATTAATTGATAAGATACCACAAGCATCATCATCATG 97127

Qy  2170 CCAGTTATGAACACAGTGCTGAAAGGATCATAGACAGGGTGGTTAAATCTGATCCCAGT 2229
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 97126 CCAGTTATGAACACAGTGCTGAAAGGATCATAGACAGGGTGGTTAAATCTGATCCCAGT 97067

Qy  2230 AGAATAAAACTTCAGTGTACCTATTTCAGGGAAGAGTTAATTCACAATTAAAACTAGTAA 2289
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 97066 AGAATAAAACTTCAGTGTACCTATTTCAGGGAAGAGTTAATTCACAATTAAAACTAGTAA 97007

Qy  2290 ATGACCAATTCTTAGGCACATTAAGTGGATTCTGAGTAAAAGAAAGGGAACAGCAGGAG 2349
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 97006 ATGACCAATTCTTAGGCACATTAAGTGGATTCTGAGTAAAAGAAAGGGAACAGCAGGAG 96947

Qy  2350 AAAGCTGTTCGCTTGGTTCGATTACCCAAATGAGCATGCTGGAACGAGCGGTTGTGAGGCT 2409
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 96946 AAAGCTGTTCGCTTGGTTCGATTACCCAAATGAGCATGCTGGAACGAGCGGTTGTGAGGCT 96887

Qy  2410 ACGTAAAAACCTCTCGCGTAGGGAGAGTACAGTGCATGAGTGTGCGGCTTTTGTCCAC 2469
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 96886 ACGTAAAAACCTCTCGCGTAGGGAGAGTACAGTGCATGAGTGTGCGGCTTTTGTCCAC 96827

Qy  2470 ACTCGTGAGGGTGAGTAATTCAGAGCCAATCACATCACAAGGATGGACACCACCTAACTC 2529
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 96826 ACTCGTGAGGGTGAGTAATTCAGAGCCAATCACATCACAAGGATGGACACCACCTAACTC 96767

Qy  2530 ATCACTTCAGGGGAGGATGAATGCTTTCATGAGAATTACACTCATAAGCATCAG 2589
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 96766 ATCACTTCAGGGGAGGATGAATGCTTTCATGAGAATTACACTCATAAGCATCAG 96707


Qy  2590 TTTGAGTAAAATTTGAGTAGAGTTAAATATGAACACATTTATACCTCTTACTAATGTCC 2649
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 96706 TTTGAGTAAAATTTGAGTAGAGTTAAATATGAACACATTTATACCTCTTACTAATGTCC 96647

Qy  2650 CACCGACACCTTTAATGTAAGACATTATTATTAAGTAGTACTTGACATTAAATGCTTA 2709
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 96646 CACCGACACCTTTAATGTAAGACATTTATTATTAAGTAGTACTTGACATTAAATGCTTA 96587

Qy  2710 TGTCTGTATATTCGTTCATCCTCGATTTTCCCAAAAGTAAGAGCATAGGAGATGAGG 2769
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 96586 TGTCTGTATATTCGTTCATCCTCGATTTTCCCAAAAGTAAGAGCATAGGAGATGAGG 96527

Qy  2770 CCTACATGCCAAGAAAACTATATTTTTACTCTTAATTCTTACTTGAGCCCAGCTTGTTG 2829
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 96526 CCTACATGCCAAGAAAACTATATTTTTACTCTTAATTCTTACTTGAGCCCAGCTTGTTG 96467

Qy  2830 TTTATCAAGTGCTTTTTGAAGACACAGCCACCCTGAATTCTTCATTCTGATACAGTGT 2889
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 96466 TTTATCAAGTGCTTTTTGAAGACACAGCCACCCTGAATTCTTCATTCTGATACAGTGT 96407

Qy  2890 CACCTTGATTTAACATTTGTAGTGTTGTTTCAAGTTACATCTCTTCATTCTTTTATA 2949
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 96406 CACCTTGATTTAACATTTGTAGTGTTGTTTCAAGTTACATCTCTTCATTCTTTTATA 96347

Qy  2950 GCAAATCAAACGTATTAGCTTCGAAATTTATCAGAGTTCATATAAAATAATTTTGCAA 3009
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 96346 GCAAATCAAACGTATTAGCTTCGAAATTTATCAGAGTTCATATAAAATAATTTTGCAA 96287

Qy  3010 AGGGTAAAAGGCTTTTTGTAATAAAATAAAAATTATTATTTCTTCTGATGAATAGA 3069
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 96286 AGGGTAAAAGGCTTTTTGTAATAAAATAAAAATTATTATTTCTTCTGATGAATAGA 96227

Qy  3070 GGCTCTTTATGCTGCTGCTAACAACCTAATTAGCTTTAAATTATCTCCTAGCAACATT 3129
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 96226 GGCTCTTTATGCTGCTGCTAACAACCTAATTAGCTTTAAATTATCTCCTAGCAACATT 96167

Qy  3130 GGTCACGTTCAATCATGCTATAGCAAAAAAAAAAA 3166
         |||||||||||||||||||||||||||||||
Db 96166 GGTCACGTTCAATCATGCTATAGCAAAAAAGAAAA 96130
```

RESULT 5
AL772303/c
```
LOCUS       AL772303              202001 bp    DNA     linear   ROD 16-SEP-2003
DEFINITION  Mouse DNA sequence from clone RP23-185P20 on chromosome 2, complete
            sequence.
ACCESSION   AL772303
VERSION     AL772303.11  GI:34787258
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 202001)
  AUTHORS   Almeida,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk  Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 16, 2003 this sequence version replaced gi:13997065.
            Sequence from the Mouse Genome Sequencing Consortium. Whole genome
            shotgun may have been used to confirm this sequence.  Sequence data
            from the whole genome shotgun alone has only been used where it has
            a phred quality of at least 30.
            ------------ Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            ------------
            During sequence assembly data is compared from overlapping clones.
```

Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compresions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-185P20 is
from the RPCI-23 Mouse BAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6.

FEATURES   Location/Qualifiers
source   1..202001
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="2"
                     /clone="RP23-185P20"
                     /clone_lib="RPCI-23"

ORIGIN

Query Match   74.6%;   Score 2361;   DB 6;   Length 202001;
Best Local Similarity 99.6%;   Pred. No. 0;
Matches 2367; Conservative   0; Mismatches   10; Indels   0; Gaps   0;

Qy   790 CATTATAACCCTGAAGTACCAGAATTGAAGGAGGAAGCACAGTTTTTGGAGGACAT 849
          |||||| ||| |||||||||||||||||||||||||||||||||||||||||||
Db 130509 CTTTTAATTCGGCAGGTACCCAGAATTGAAGGAGGAAGCACAGTTTTTGGAGGACAT 130450

Qy   850 CTCCACCTATGAGATGCATTCCTCCTCCTGCCCAGCATTTCCTATCGGGCAACACAGG 909
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 130449 CTCCACCTATGAGATGCATTCCTCCTCCTGCCCAGCATTTCCTATCGGGCAACACAGG 130390

Qy   910 CATCTCTTTAAAGTCTACCAAACACTCAAAGAGTCAAAATGAGGCAAAAGGTTCTT 969
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 130389 CATCTCTTTAAAGTCTACCAAACACTCAAAGAGTCAAAATGAGGCAAAAGGTTCTT 130330

Qy   970 CTTCCATCCCAGGTACCTGAGACACCTCGCTCTTTCTGGAGAACTAAAGGGGTGACTGC 1029
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 130329 CTTCCATCCCAGGTACCTGAGACACCTCGCTCTTTCTGGAGAACTAAAGGGGTGACTGC 130270

Qy   1030 ATACCGCTTGTCCACAGGCTTGATGATTGCAAGTGTCGCTGTGGAACGTGTGAAAACGT 1089
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 130269 ATACCGCTTGTCCACAGGCTTGATGATTGCAAGTGTCGCTGTGGAACGTGTGAAAACGT 130210

Qy   1090 GAAGCTCTACGGATTCTGGCCTTTCTCTAAGACTATCGAAGAGCTATCGAAGAGCAGTGCTCA 1149
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 130209 GAAGCTCTACGGATTCTGGCCTTTCTCTAAGACTATCGAAGACACCCCACTCAGTCACCA 130150

Qy   1150 CTACTATGATAACATGTTACCTAAGCATGGTTTCCACCAGTGCCTAAAGAATACAGCCA 1209
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 130149 CTACTATGATAACATGTTACCTAAGCATGGTTTCCACCAGTGCCTAAAGAATACAGCCA 130090

Qy   1210 AATGCTCCAGCTCCATATGAGAGGAATCCTCAAACTGCAATTCAGCAAATGTGAAACGGC 1269
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 130089 AATGCTCCAGCTCCATATGAGAGGAATCCTCAAACTGCAATTCAGCAAATGTGAAACGGC 130030

Qy   1270 TTAAACGTTTCTTAGAAGGAGAATAATTTCAGGAGGTGGATGGTGGATGGTGTCACGATCTC 1329
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 130029 TTAAACGTTTCTTAGAAGGAGAATAATTTCAGGAGGTGGATGGTGGATGGTGTCACGATCTC 129970

Qy   1330 CAAAAGCCAATAGAGAAGGCACAGAGAAAGCATGAATTCAGAAAGGCGCTCTCCCACTT 1389

Db 129969 CAAAAGCCAATAGAGAAGGCACAGAGAAAGCATGAATTCAGAAAGGCGCTCTCCCACTT 129910

Qy   1390 GTCTAGACCAAAGCCACCCGCCCCACTCACTTTGCAGCTCCAGAGTCACTCATTCTC 1449
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 129909 GTCTAGACCAAAGCCACCCGCCCCACTCACTTTGCAGCTCCAGAGTCACTCATTCTC 129850

Qy   1450 ACCTTCAACGTTCTTCCTGAGCATAGAGACCAAACATCAGACTTGATAAGTAAGAAAT 1509
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 129849 ACCTTCAACGTTCTTCCTGACATAGAGACCAAACATCAGACTTGATAAGTAAGAAAT 129790

Qy   1510 GAGATAATTTCAAATCATCAGAATTGATTGAGCCAGGGTCTCCAGAATGCTTC 1569
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 129789 GAGATAATTTTCAAATCATCAGAATTGATTGAGCCAGGGTCTCCAGAATGCTTC 129730

Qy   1570 CTTGTTCCTATCCATGATAGCCTCCCACCTTTATCAGATGGTAATGAAACTGTGCAA 1629
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 129729 CTTGTTCCTATCCATGATAGCCTCCCACCTTTATCAGATGGTAATGAAACTGTGCAA 129670

Qy   1630 TTGTGCCAAAGACCCTTTCTGAGAGATGTCTGAATCATGCGCCGAGTTTTACACACA 1689
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 129669 TTGTGCCAAAGACCCTTTCTGAGAGATGTCTGAATCATGCGCCGAGTTTTACACACA 129610

Qy   1690 GCTCTTCCTTATAAATAAATCCTCCATTCCTCCCCTCCTCCTAGTAGAGTACAGAAACAAA 1749
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 129609 GCTCTTCCTTATAAATAAATCCTCCATTCCTCCCCTCCTCCTAGTAGAGTACAGAAACAAA 129550

Qy   1750 TACCCTTGATGATTCAGGAAGAAGTCTTTTTACTTAGCAATGTGCCGCTTCTGATT 1809
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 129549 TACCCTTGATGATTCAGGAAGAAGTCTTTTTACTTAGCAATGTGCCGCTTCTGATT 129490

Qy   1810 CAGTTCGCTTGTGACATTAAGCGGGTTGGGGTTTGGGGTTTGGGTTGGATTGGGGCGTTTCTTCA 1869
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 129489 CAGTTCGCTTGTGACATTAAGCGGGTTGGGGTTTGGGGTTTGGGTTGGATTGGGGGCGTTTCTTCA 129430

Qy   1870 CTTCTTTGTCTATATTTCCTACCTTTATCAGTTTGTATTGGAGCTTCCTGCTTGGG 1929
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 129429 CTTCTTTGTCTATATTTCCTACCTTTATCAGTTTGTATTGGAGCTTCCTGCTTGGG 129370

Qy   1930 ATTCTGCAATTCTCTCCCACGACAGGATCAACTCAATGACATAAAGTAGTTCAAACA 1989
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 129369 ATTCTGCAATTCTCTCCCACGACAGGATCAACTCAATGACATAAAGTAGTTCAAACA 129310

Qy   1990 TCCATTGCTTCTCACATGTTTTTCCATAAAGTTACTCATCTGATTTTATTTAAAATAGT 2049
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 129309 TCCATTGCTTCTCACATGTTTTTCCATAAAGTTACTCATCTGATTTTATTTAAAATAGT 129250

Qy   2050 GAACATCCACTTGATATCAGACCGAGGACCATCCTCCATTGGAGAATATGGAAGATATTG 2109
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 129249 GAACATCCACTTGATATCAGACCGAGGACCATCCTCCATTGGAGAATATGGAAGATATTG 129190

Qy   2110 TCACTGGCAGAAAAGCAGGTGTGGCCATTAATTGATAAGATACCACAAGCATCATCATG 2169
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 129189 TCACTGGCAGAAAAGCAGGTGTGGCCATTAATTGATAAGATACCACAAGCATCATCATG 129130

Qy   2170 CCAGTTATGAACACAGTGCTGAAGGATCATAGACACAGTTCGGTGTTAAATCTGATCCCAGT 2229
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 129129 CCAGTTATGAACACAGTGCTGAAGGATCATAGACACAGTTCGGTGTTAAATCTGATCCCAGT 129070

Qy   2230 AGAATAAACTTCAGTGTACCTATCAGGGAAGAGTTAATTTCACAATTAAAACTAGTAA 2289
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 129069 AGAATAAACTTCAGTGTACCTATTCAGGGAAGAGTTAATTTCACAATTAAAACTAGTAA 129010

Qy   2290 ATGAACCAATTCTTAGGCACATCAGTGGATTCGAGTAAAGAAAGGGAACAGCAGGAG 2349
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 129009 ATGAACCAATTCTTAGGCACATCAGTGGATTCGAGTAAAGAAAGGGAACAGCAGGAG 128950

Qy   2350 AAAGCTGTTCGCTTGGTTCTGAACCCAAATGAGCATGCTGAAGGAGGTTGTGAGGCT 2409
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 128949 AAAGCTGTTCGCTTGGTTCTGACCCCAAATGAGCATGCTGAAGGAGGTTGTGAGGCT 128890

Qy   2410 ACGCTAAAACCTCTGCGTAGGGAGAGTACAGTGCATGAGTGCGCGGCTTTTGTCCAC 2469

Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
Db  128889 ACGGTAAAACCTCGCTAGGGAGAGAGTACAGTGCATGAGTGTGGCGGCTTTTGTCCAC 128830
Qy    2470 ACTCGTGAAGGGTGAGTAATTCAGAGCCAATCACAAGGATGGACACACCTAACTC 2529
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  128829 ACTCGTGAAGGGTGAGTAATTCAGAGCCAATCACAAGGATGGACACACCTAACTC 128770
Qy    2530 ATCACTTCAGGGGGAGGATGAATGCTTCATGAGAAATTACACTCATAAGCTAAGCATCAG 2589
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  128769 ATCACTTCAGGGGGAGGATGAATGCTTCATGAGAAATTACACTCATAAGCTAAGCATCAG 128710
Qy    2590 TTTTGAGTAAAATTGAGTAGATGTTAAATATGAACATTTATACCTCTTACTAAATGTCC 2649
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  128709 TTTTGAGTAAAATTGAGTAGATGTTAAATATGAACATTTATACCTCTTACTAAATGTCC 128650
Qy    2650 CACCGACACCTTTAATGTAAGCACATTTATTATTAAGTACTTGACATTAAATGCTTA 2709
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  128649 CACCGACACCTTTAATGTAAGCACATTTATTATTAAGTACTTGACATTAAATGCTTA 128590
Qy    2710 TGTCTGTATATTCTGTTCATCCATCGATTTCCCAAAAAGTAAGAGCATAGGAGATGAGG 2769
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  128589 TGTCTGTATATTCTGTTCATCCATCGATTTCCCAAAAAGTAAGAGCATAGGAGATGAGG 128530
Qy    2770 CCTACACATGCCAAGAAAACTATAAATTTTACTCTTTAATTCTTACTTGAGCCAGCTTGTTG 2829
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  128529 CCTACACATGCCAAGAAAACTATAAATTTTACTCTTTAATTCTTACTTGAGCCAGCTTGTTG 128470
Qy    2830 TTTATCAAGTGCTTTTTGAAGAGACAGCACCCTGTGAATTCTTCATTCTGATACAGTGT 2889
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  128469 TTTATCAAGTGCTTTTTGAAGAGACAGCACCCTGTGAATTCTTCATTCTGATACAGTGT 128410
Qy    2890 CACCTTGTATTAACATTTGTAATGTGTTTCAAGTTACACATCTCTTTTCATTCTTTTATA 2949
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  128409 CACCTTGTATTAACATTTGTAATGTGTTTCAAGTTACACATCTCTTTTCATTCTTTTATA 128350
Qy    2950 GCAAATCAAACGTATTAGCTTCAGAAATTTATCAGAGTTCATATATAAATATTTTGCAA 3009
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  128349 GCAAATCAAACGTATTAGCTTCAGAAATTTATCAGAGTTCATATATAAATATTTTGCAA 128290
Qy    3010 AGGGTAAAAGGCTTTTTTGTTTAAATAAATAAAATTTATTATTTTCTTCTGATGAATAGA 3069
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  128289 AGGGTAAAAGGCTTTTTTGTTTAAATAAATAAAATTTATTATTTTCTTCTGATGAATAGA 128230
Qy    3070 GGCTCTTTTATGCTGCTAATGAACCTAATGAGCTTAAATTATCTCCCTAGCAACATT 3129
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  128229 GGCTCTTTTATGCTGCTAATGAACCTAATGAGCTTAAATTATCTCCCTAGCAACATT 128170
Qy    3130 GGTCACGTTCAATCATGCTATTAGCAACAAAGAAAAAAA 3166
           |||||||||||||||||||||||||||||||||||
Db  128169 GGTCACGTTCAATCATGCTATTAGCAACAAAGAAA 128133
```

REFERENCE AUTHORS Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Dramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,C.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Santa,G., Garcia,A., Garner,T., Garza,M.,
Gebregeorgis,E., Geer,K., Gill,R., Grady,A., Guerra,N., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hales,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., ladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., ulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Ji,kly,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., kly,S., Khan,Z., King,L., Kovar,J.,
Kowis,C., Kraft,C.L., Leow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., ondon,P. Longacre,S., Lopez,J.,
Lorensuhewa,L., Loulsege,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindart,e.M., Mahmoud,R., Martin,R., Martinez,B.,
Mangum,B., Mapua.P., Martin,K., Martin,R., Martinez,B.,
Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner, Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Norris,S.,
Nwaokelemeh,O. Okwuonu, Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
Plopper,P., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puazo,M., Quiroz,J., Raclin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Roj,A., Rose,M., Rose,R., Rui,S.J.,
Sanders,W., Savery,G., Sherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., ong,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,R., stton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,V., Vera,V., Villana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willison,R., wlczyk,K., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,Z., Yakub,S., Yen,H., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss, Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.

TITLE   Direct Submission
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 236283)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (09-MAY-2002) Human Genome Sequencing Center, Department
           of Molecular and Human Genetics, Baylor College of Medicine, One
           Baylor Plaza, Houston, T 77030, USA
REFERENCE 3 (bases 1 to 236283)
AUTHORS   Rat Genome Sequencing Consortium.
TITLE   Direct Submission
JOURNAL   Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
           of Molecular and Human Genetics, Baylor College of Medicine, One
           Baylor Plaza, Houston, T 77030, USA
COMMENT   On Nov 19, 2002 this sequence version replaced gi:23264993.
           The sequence in this assembly is a combination of BAC based reads
           and whole genome shotgun sequencing reads assembled using Atlas
           (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
           in the feature table below represents a scaffold in the Atlas
           assembly (a 'contig-scaffold'). Within each contig-scaffold,
           individual sequence contigs are ordered and oriented, and separated
           by sized gaps filled with Ns to the estimated size. The sequence
           may extend beyond the end of the clone and there may be unordered
           contigs within a contig-scaffold that consist entirely of whole
           genome shotgun sequence reads. Both end sequences and whole genome
           shotgun sequence only contigs will be indicated in the feature
           table.
           -------- Genome Center
           Center: Baylor College of Medicine
           Center code: BCM
           Web site: http://www.hgsc.bcm.tmc.edu/
           Contact: hgsc-help@bcm.tmc.edu

RESULT 6
AC120773
LOCUS   AC120773   236283 bp   DNA   linear   HTG 19-NOV-2002
DEFINITION Rattus norvegicus clone CH230-65J6, WORKING DRAFT SEQUENCE, 2
            unordered pieces.
ACCESSION   AC120773
VERSION   AC120773.5 GI:25082599
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE   Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1 (bases 1 to 236283)
AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,

```
--------- Project Information
       Center project name: GZJM
       Center clone name: CH230-65J6
--------- Summary Statistics
       Assembly program: Phrap; version 0.990329
       Consensus quality: 221191 bases at least Q40
       Consensus quality: 224201 bases at least Q30
       Consensus quality: 225744 bases at least Q20
       Estimated insert size: 228488; sum-of-contigs estimation
       Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
* NOTE: Estimated insert size may differ from sequence length
*      (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
*      consists of 2 contigs. The true order of the pieces
*      is not known and their order in this sequence record is
*      arbitrary. Gaps between the contigs are represented as
*      runs of N, but the exact sizes of the gaps are unknown.
*      This record will be updated with the finished sequence
*      as soon as it is available and the accession number will
*      be preserved.
*          1    224842: contig of 224842 bp in length
*     224843    224942: gap of unknown length
*     224943    236283: contig of 11341 bp in length
FEATURES             Location/Qualifiers
     source          1..236283
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-65J6"
     misc_feature    37445..38277
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end sequence:BH351867"
     gap             224843..224942
                     /estimated_length=unknown
ORIGIN

Query Match      44.5%;  Score 1408.4;  DB 12;  Length 236283;
Best Local Similarity   80.4%;  Pred. No. 0;
Matches 1945;  Conservative   0;  Mismatches   391;  Indels   82;  Gaps   22;

Qy    806 TACCAGAATTTGAAGGAGGAAGAAAGCACAGTTTTGGAGGACATCTCCACCTATGGAGAT 865
          ||||| |||||||||||||||||||||||| |||||  ||||   ||||||||  |
Db  57567 TACAAGAATTTGAAGGAGGAAGAAAGCACGGTTTTCCGGCGGATATTTCCGGCCATGGGGAC 57626

Qy    866 GCATTCCCTCCTGCCAGCATTTCCTATCGGGCCAACACAGCATCTCTTTTAAAGTC 925
          ||||||||| ||||||| ||| || |||||||||    ||||||||||||||||||
Db  57627 GCATTCCCTTTGCCAGCCTTTGCCAGCCTTTCCTATCGGGCCAGCATCTCTTTTAAAGTC 57686

Qy    926 TACCAAACACTCAAAGAGTCAAAAATGAGGCAAAAATGATGTGGGGTGACTGCATACCG 985
          ||||||||||||||||||||  | |||||| ||||||  | | |||||  ||||||||
Db  57687 TACCAAACACTCAAAGAGTCAAAAGAGTGAGGCAAAAGGTCTAAAGGTTCTCTTCTTCCACCCCCAGGTAC 57746

Qy    986 CTGAGACACCTCGCTCTTTTCTGGAGGACTAAAGGGGTGACTGCATACCGCTTGTCCACA 1045
          |||||||||||||||||||| |||||||||||||  |   || ||| || | |||||
Db  57747 CTGAGACACCTCGCTCTTTTCTGGAGGACTAAAGGGGTGAGGACTAAAGGTTCTCTTCTTCCACCCCCAGGTAC 57806

Qy   1046 GGCTTGATGATTGCAAGTGTCGCTGTGGAACTGTGTGAAAAGTGAAGCTCACGGATTC 1105
          |||||||||  ||| ||| | | || | ||||||||||||| ||| || || |||||
Db  57807 GGCTTGATGATCGCGAGTGGAGCATCGCGGTGGAACTGTGTGAAAATGTCACGCTCACGGATTC 57866

Qy   1106 TGGCCTTTCTCTAAGACTATGCAGAGACACCCCACTCAGTCACCACTACTATGATAACATG 1165
          |||| ||||||||||||||| | |||||| |||  |||  |||||||||||||||||
Db  57867 TGGCCGTTCTCTAAGACCGTCTAAGACCGTGGAGAAATTCCGTCGAGCCACCACTACTATGATGACAACAG 57926

Qy   1166 TTACCTAAGCATGGTTCCACCAGATGCCTAAAGAATACAGCCAAATGCTCCAGCTCCAT 1225
          |||||||| ||   |||||||| || ||||| ||||||||| ||||||||||||||||
Db  57927 CTACCTAAGCACCGTTTCCACCAGATGCCTAAGGAATACAGCCAAATGCTCCAGCTCCAT 57986

Qy   1226 ATGAGAGAATCCTCAAACTGCAATTGCAGCAAATGTGAAACGGCTTAACGTTTCTTAGAA 1285
          ||||||| |||||||||||| | ||||  | |  || |||  |||||||  |||||  |
Db  59057 TTTCACAATTAAAACTAGAAAG-GCAAGTTAAGTGAATTGGATTCGGGGTA 59115
```

```
Qy   2329 AAAGAAAG------GGAACAGCAGGAGAAAGCTGTTCGCTTGGTTCTGATTACCCAAATGA 2383
          |||| |||      ||||||| || || |  | |||| ||| | |||  ||| |||||||
Db  59116 AAAAAAAAAAATGGAACAAGAGAGGAGGAGGCGGTTTCCTCGACTATGATTTTCCAAATGA 59175

Qy   2384 GCATGCTGGAAGGAGGTTGTGAGGCTACGCTAAAACCTCGCCTAGGGAGAGAGTACAGT 2443
          | |||||| || || |||||||| |  | ||| ||| |  ||| ||||||||||||
Db  59176 GTATGCTGATAGTAGACTGTGAGGTTGCCTGACACACTCGTGTTAGGAGAGAGTACAGG 59235

Qy   2444 GCATGAGTGTGGGCGGC-TTTGTCCACTCGT---GAAGGGTGAGTAATTCAGAGCCAA 2499
          ||||| | ||| |  |  |||||||| | | |   ||| |||||||||||||||||||
Db  59236 GCATGGGGGTGGTAGCTTTTTGTCCACACTGGTGAAGAAGTGAGTAATTCAGAGCCAA 59295

Qy   2500 TCACATCACA---------AGGATGGACACCTAACTCATCACTTCAGGGGGAGATGA 2549
          |||| |||||         ||||| || ||  | |   | ||| |||| | ||||||
Db  59296 TCACTTCACAAGGAGGAGGCGGTTAGGATAGCAGCCGCCTGACCATTACTCCAGGGAAAGATGA 59355

Qy   2550 ATGCTTCATGAGAAAATTACACTCATAAGCTAAGCACATCAGTTTGAGTAAAATTTGAGTA 2609
          ||| ||||||||||||||| | |||||||| ||| | |||||| | | ||||||| |||||
Db  59356 ACGTTTTCATGAAAAATTGCACTCATCGGCTAAGCATCAGTCTTGAGTAAAATTTGAATA 59415

Qy   2610 GATGTTAAATATGAACATTTATTACCTCCTACATATGTCCCACCGACACCTTTAATGTA 2669
          |||||||||||| ||||||||| | | |||||| ||| ||  | | || || ||||| |
Db  59416 GATGTTAAATATGAGACAATTTACACCCCTTAC----GTCCCAGGGATACCCCTTAATGTA 59471

Qy   2670 AGCACATTTATTTATTTAAGTTACTTGACATTAAATGCTTATGTCTGTATATTC-TGTTCA 2728
          ||| ||||| ||| ||||| | ||| ||  |  | |||  |||  ||| |||| | ||| |
Db  59472 AGCACATTTCTTAGGTTATTTGACT--TATTACATATTTATGTGTGTACATTCTTGTTCA 59529

Qy   2729 TCCATCGATTTTCCCAAAAGTAAGAGCATAGGAGATGAGGCCTACATGCCAAGAAAACT 2788
          ||| |||| ||||||| ||| ||| |  ||| |||| || |||  | |||||||||||||
Db  59530 TCCACGGATGGTCCCCAGAACTAGGAGTCGGGATCGCGGTCCACTTGCCAAGAAAACT 59589

Qy   2789 ATAAATTTACTCTTAATTCTTACTTGAGCGGCCAGCTTGTTGTTTATCAAGTGCTTTTTTG 2848
          ||||||| | |  |  ||||| || ||| ||||| ||| || |||||  |||| || |||
Db  59590 ATAAATTTACTCTTACCTCTCTCACCTGAGCCGCCATTCCA-CTGATACTTTCCTG 59648

Qy   2849 AAGAGACAGCACCCTGTGAATTCTTCATTCTGATACAGTGTCACCTTGTATTTAACATTT 2908
          || | ||| |  ||||||||||||||||||| | | || |||||||| |  ||| |||||
Db  59649 AAGGAAGATTACCCTGTGAATTCTTCATTCTGATACAGTGTCACCTTGTTATTTGACATTT 59708

Qy   2909 GTAATGTTGTTTCAAGTTACATCTCTTTCATTCTTTATAGCAAATCAAAGTATTAGC 2968
          |||  |  |       ||||||| | |||| | |||||| || ||| || |||||||
Db  59709 GTAA---CGTGTCAAGTTGCTTCTCTCCTTTCGTTTTATAGCCAATCAAACTTGTTAGC 59765

Qy   2969 TTCAGAAATTTATCAGAAGTTCATATATAAATTTGCAAAGGGTAAAAGGCTTTTTTG 3028
          |||||||  ||| ||||||||||||| | | |||||||||| ||| ||||||||||
Db  59766 TTTAGAAATATATCAGAAGTTCATATATAAATATTTTGCAAAGGGTAAAAAGGCTTTTTG 59825

Qy   3029 TTAAAATAAA-----ATTTATTATTTTCTTCGATGAATAGAGGCTCTTTTATGCTG 3084
          |||| ||| ||     ||| ||| |||| ||| | ||| |||| |||||||||||||
Db  59826 TTAAATAAATCAAATGTATTATTAGCTTATTATCTCCTAGCAACATTGGTCACGTTTCAATC 59885

Qy   3085 CTGCTAATGAACCTAATTAGCTTTAATTATTATCTCCTAGCAACATTGGTCACGTTTCAATC 3144
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  59886 CTGCTAATGGGTTAATTAGCTTTAAATAATCTCCTAGCAACATTGGTCATGTTTCAATC 59945

Qy   3145 ATGCTATTAGCAAAAAAAA 3162
          ||||||||||  ||| |||
Db  59946 CCACTATTTGCAAAGAA 59963
```

REFERENCE   1 (bases 1 to 264245)

AUTHORS   Muzny,D.Marie. Metzker,... Lee, Abramzon,S., Adams,C., Alder,J.,
          Allen,C., Allen,H., Aleb...cooks,S., Amin,A., Anguiano,D.,
          Anyalebechi,V., Aoyagi,A...Ayodeji,M., Baca,E., Baden,H.,
          Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
          Biswalo,K., Blair,J., Bl...xenburg,K., Blyth,P., Brown,M.,
          Bryant,N., Buhay,C., Bur...rrell,K., Calderon,E.,
          Cardenas,V., Carter,K....avazos,I., Ceasar,A. Center,A.,
          Chacko,J., Chavez,D.,Ch...G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
          Cleveland,C., Cockrell,R...Cox,C., Coyle,M., Cree,A., D'Souza,L.,
          Davila,M.L., Davis,C.,...vy-Carroll,L., De Anda,C., Dederich,D.,
          Delgado,O., Denson,S.,D...drano,C., Ding,Y., Dinh,H., Divya,K.,
          Draper,H., Dugan-Rocha,S...Dunn,A., Durbin,K., Duval,B., Eaves,K.,
          Egan,A., Escotto,M. Eug...ine,C., Evans,C.A., Falls,T., Fan,G.,
          Fernandez,S., Finley,M.,...Flagg,N., Forbes,L., Foster,M., Foster,P.,
          Fraser,C.M., Gabisi,A....anta,R., Garcia,A., Garner,T., Garza,M.,
          Gebregeorgis,E., Geer,K....Gill,R., Grady,M., Grady,W., Guevara,W.,
          Gunaratne,P., Haaland,W...Hamil,C., Hamilton,C., Hamilton,K.,
          Harvey,Y., Havlak,P., Ha...les,A., Henderson,N., Hernandez,J.,
          Hernandez,R., Hines,S....ladun,S.L. Hodgson,A., Hogues,M.,
          Hollins,B., Howells,S...ulyk,S. Hume,J., Idlebird,D., Jackson,A.,
          Jackson,L., Jacob,L., Ji...ng,H., Johnson,B., Johnson,R., Jolivet,A.,
          Karpathy,S., Kelly,S.,K...ng,L., Khan,Z., King,L., Kovar,C.,
          Kowis,C., Kraft,C.L., Le...ow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
          Liu,J., Liu,W., Liu,Y...ondon,P., Longacre,S., Lopez,J.,
          Lorensuhewa,L., Loulsege...Lozado,R., Lu,X., Ma,J.,
          Maheshwari,M., Mahindart...M., Mahmoud,M., Malloy,K., Mangum,A.,
          Mangum,B., Mapua,P., Mar...Martin,K., Martin,R., Martinez,E.,
          Mawhiney,S., McLeod,M.P...McNeill,T.Z., Meenen,E.,
          Milosavljevic,A., Miner,...Minja,E., Montemayor,J., Moore,S.,
          Morgan,M., Morris,K., Mo...wton,N., Nguyen,N., Norris,S.,
          Nankervis,C., Neal,D., N...
          Nwaokelemeh,O., Okwuonu,...Olarrpunsagoon,A., Pal,S., Parks,K.,
          Pasternak,S., Paul,H., P...rez,A., Perez,L., Pfannkoch,C.,
          Plopper,F., Poindexter,A...Popovic,D., Primus,E., Pu,L.L.,
          Puazo,M., Quiroz,J., Rac...lin,E., Reeves,K., Regier,M.A., Reigh,R.,
          Reilly,C., Reilly,M., Re...Y., Reuter,M., Richards,S., Riggs,F.,
          Rives,C., Rodkey,T., Roj...,A., Rose,M., Rose,R., Ruiz,S.J.,
          Sanders,W., Savery,G.,S...herer,S., Scott,G., Shatsman,S., Shen,H.,
          Shetty,J., Shvartsbeyn,A...Sisson,I., Sitter,C.D., Smajs,D.,
          Sneed,A., Sodergren,E.,S...ong,X.-Z., Sorelle,R., Sosa,J.,
          Steimle,M., Strong,R., S...ton,A., Svatek,A., Tabor,P., Taylor,C.,
          Taylor,T., Thomas,N., Th...mas,S., Tingey,A., Trejos,Z., Usmani,K.,
          Valas,R., Vera,V., Villa...ana,D., Waldron,L., Walker,B., Wang,J.,
          Wang,Q., Wang,S., Warren...J., Warren,R., Wei,X., White,F.,
          Williams,G., Willson,R...Wleczyk,R., Wooden,H., Worley,K.,
          Wright,D., Wright,R., Wu...J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
          Yu,F., Zhang,J., Zhou,J...Zhou,X., Zhao,S., Dunn,D., von
          Niederhausern,A., Weiss,...Smith,D.R., Holt,R.A., Smith,H.O.,
          Weinstock,G. and Gibbs,R.A.

TITLE   Direct Submission

JOURNAL   Unpublished

REFERENCE   2 (bases 1 to 264245)

AUTHORS   Worley,K.C.

TITLE   Direct Submission

JOURNAL   Submitted (28-SEP-2001) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA

REFERENCE   3 (bases 1 to 264245)

AUTHORS   Rat Genome Sequencing Consortium.

TITLE   Direct Submission

JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA

COMMENT   On May 10, 2003 this sequence version replaced gi:24819385.
          The sequence in this assembly is a combination of BAC based reads
          and whole genome shotgun sequencing reads assembled using Atlas
          (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
          in the feature table below represents a scaffold in the Atlas
          assembly (a 'contig-scaffold'). Within each contig-scaffold,
          individual sequence contigs are ordered and oriented, and separated

RESULT 7

AC096804/c

LOCUS   AC096804   264245 bp   DNA   linear   HTG 10-MAY-2003

DEFINITION Rattus norvegicus clone CH230-44K10, *** SEQUENCING IN PROGRESS
            ***, 2 unordered pieces.

ACCESSION   AC096804

VERSION   AC096804.6 GI:30521409

KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.

SOURCE   Rattus norvegicus (Norway rat)

ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.

by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

-------- Genome Center
    Center: Baylor College of Medicine
    Center code: BCM
    Web site: http://www.hgsc.bcm.tmc.edu/
    Contact: hgsc-help@bcm.tmc.edu
------- Project Information
    Center project name: GFFQ
    Center clone name: CH230-44K10
------ Summary Statistics
    Assembly program: Atlas 3.0;
    Consensus quality: 215957 bases at least Q40
    Consensus quality: 221425 bases at least Q30
    Consensus quality: 225088 bases at least Q20
    Estimated insert size: 225712; sum-of-contigs estimation
    Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length.
*   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
*   consists of 2 contigs. The true order of the pieces
*   is not known and their order in this sequence record is
*   arbitrary. Gaps between the contigs are represented as
*   runs of N, but the exact sizes of the gaps are unknown.
*   This record will be updated with the finished sequence
*   as soon as it is available and the accession number will
*   be preserved.
*   1   262878: contig of 262878 bp in length
*   262879   262978: gap of unknown length
*   262979   264245: contig of 1267 bp in length.

FEATURES   Location/Qualifiers
     source   1..264245
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-44K10"
     misc_feature   1..1967
                     /note="wgs end_extension
                     clone_end:T7"
     misc_feature   complement(3314..4075)
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH294791"
     misc_feature   80231..81276
                     /note="wgs contig"
     misc_feature   complement(260505..261297)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH294793"
     gap   262879..262978
                     /estimated_length=unknown
ORIGIN

Query Match   44.5%; Score 1408.4; DB 12; Length 264245;
Best Local Similarity 80.4%; Pred. No: 0;
Matches 1945; Conservative 0; Mismatches 391; Indels 82; Gaps 22;

Db   247382 TACCAAACACTCAAAGAGTCTA AGTGAGGCAAAAGGTTCTCTTCTTCCACCCCAGGTAC 247323
                                    |||||||||||||||||||||||||||||||||||||
Qy   986 CTGAGACACCTCGGCTCTTTTCT GAGAACTAAAGGGGTGACTGCATACCGCTTGTCCACA 1045
                                    |||||||||||||||||||||| |||||| |||||||
Db   247322 CTGAGACACCTCGGCTCTTTTCT GAGAACTAAAGGGGTGACTGCGTACCGCTTGTCCACG 247263

Qy   1046 GGCTTGATGATTGCAAGTGCG GTGTGGAACTGTGTGAAAACGTGGAAGCTCTACGGATTC 1105
                                  ||||| |||||||| |||||||||||||||||||||||||
Db   247262 GGCTTGATGATGCGGACGATCG GTGGAACTGTGTGAAAATGTCACGGCTCTACGGATTC 247203

Qy   1106 TGGCCTTTCTCTAAGACTATCG GAGACACCCCAGTCAGTCACCACTACTATGATAACATG 1165
                                  |||| ||||||||| ||||||||||||||||||||
Db   247202 TGGCCGTTCTCTAAGACCGTCG GAGAAATTCCCTGAGCCACCACTACTATGACAACAAG 247143

Qy   1166 TTACCTAAGCATGTTCCACC GATGCTAAAGAATACAGCCAAATGCTCCAGCTCCAT 1225
                                  |||||||| |||||||||||||||||||||||||||||||
Db   247142 CTACCTAAGCACGGTTCCACC CATGCCTAAGGAATACAGCCAAATGCTCCAGCTCCAT 247083

Qy   1226 ATGAGAGAAATCCTCAAACTGC ATTCAGAAATGTGAAACGGCTTAACGTTCTTAGAA-- 1285
                                  |||| ||| |||||||||||| |||||||||||||||||||||
Db   247082 ATGAAAGGGATCCTCAAACTGC GTTCAGAAATGTGAAGCCCGTTAGCGTTTCTTAAAG 247023

Qy   1286 GGAGAATAATTTCAGGAGTGG GTGGAATGTGTCACAGCCATCCCAAAAAGCCAATAGAA 1345
                                  |||||||| ||| |||||||| || |||||||||||||||||
Db   247022 TGCGAACCATTTCAGTAGGTGG GTGGGTGCATCGCAGTGTCTCC-AAACGTGATGGAG 246964

Qy   1346 GAAGGCACAGAGAAGCATGAA TACAAAGGCGCTCTCCACTTGTCTAGACCAAAGC-- 1403
                                  ||||||||||||| ||| ||| || |||||||| |||||||||||
Db   246963 AAAGGCACAGAACATGGA-GGA TGCACAAGAGCTCCCGACTTGGCTAGACCAAAGCCC 246905

Qy   1404 -CACCCGCCCCCACTCACTTTG GAGCTCCAGAGTCACTCATTCTCACCTTCAACGTTC 1462
                                  |||| ||||| ||||||||||| ||||||||||||||||| |||||||
Db   246904 GCCCCGCCCGCCCCCGCTCACTTTG AGCCCCACGAGTCACTCACTTCTCAACTTCAATGTTC 246845

Qy   1463 TTTCTCTGAGAATAGAGACCAA CATCAGACTTGGATAAGTAAGTAAAATGAGATAATTTTC 1522
                                  ||||||||||||| ||| ||| ||| |||| ||||||||||| |||||
Db   246844 TTTCTCTGAGAATGGAGGTCAC TATCAGACTCCGACAAGTAAAATGAGATCATGTTTA 246785

Qy   1523 AAATCATCATAGAATTTGATT AGCCAGGGTCTCTCAGAATGTTCCTTGTTCCTATCC 1582
                                  |||| ||||||||||| |||| |||||||||||||||| ||||||||| |||||
Db   246784 AATCGGCCATAGAGATTTGATC AGGCAGCGACTCGGAGGAGTGCTTTCCGTTCCTATTC 246725

Qy   1583 ATGATAGCCATTCCCACCTTTA CAGAGTGGTAATGAAACTGTGCAATGTGCCAAAGAC 1642
                                  ||| ||||||||||| ||||| |||||||| |||||| |||||||||||||||
Db   246724 GTGGTAGCCATTCCCGCCATTA CAGAGTGGTGACGAAACTGTGCAACGTGTGCCAAAGAG 246665

Qy   1643 CCTTTCTGAAGAGAATGTCTGA TCA------TGGCGGCGAGTTTTACACAGCTCTC 1696
                                  ||||||||||||||||||| || |||||||||||||||||||||||
Db   246664 CCCTTCTGAAGAGAATGTCCGC CATGAGGGTGCACCGAGCTCCCCTCACAGCCCTTC 246605

Qy   1697 CTTTATAAATAA--ATCCTTC ATTCTCCCTCCTCCAGAGTACAGAAACAAAATACC 1753
                                  ||||| |||||| |||| |||||||||||||| ||||||||| |||||||
Db   246604 CTTTATCAATAACCTTTCCACTT CCTCCCCCCAAATGAGTACCGAAACAAAACACC 246545

Qy   1754 CTTGATGATTCAGGAGGAAAAG CTTTTTTTTACTTAGCAATGTGCCTGCTTCTGATTCAGT 1813
                                  |||||||||||||||||| ||| ||||||||||||||||||| |||||||| |||
Db   246544 CTTGATGATTCAGGAGGAGAAAC CTTTTCTACTTAGCAATGTGCCTGCTTCTGACTCGGT 246485

Qy   1814 TCGCTTGTGACATTAAGCTGGG TGGGGTTTGGTTTGGTGATTGGGTTTCTTCACTTC 1873
                                  |||| ||||||| ||||| ||| |||||||||||| |||||||||||||| |||| |||
Db   246484 TCACTTGTGA-TGTAATCTGCGG TGAGGTGTTGGTGGGGATTTTGGGGCGTTTCCTCGCCTC 246426

Qy   1874 TTTTGTCTATATTTCCTTACC TATCAGTTTGTATTCGCTTCCGTTCTTCTGGGATTC 1933
                                  || ||| |||||||||| |||| ||||| ||| ||||||| |||||| ||||||||| ||
Db   246425 TTATCAGTTTGTATTCGCTACC CTATCAGTTTGTCATACTTTGCCTACCCTACCTTCCTTCTCTGGGGTTC 246366

Qy   1934 TGCAAT------- 1976
                                  ||||||
Db   246365 TCTTCTTTGGGGTTCTGCTCCCCCCACACACCGGACCAGGCATCCAATGACACACTCTTTTAAAGTC 246306

Qy   1977 AGTAGTTCAAACATCCATTGCT TCACATGTT----TTATCCATAAAGTTACTCATCTGA 2033
                                  ||||| ||||||||||||||| | ||
Db   806 TACCAGAATTTGAAGGAGGAAGAAGACAGTTTTGAGGACATCTCCACCTATGGAGAT 865
              ||||||||||||||||||| |||||| ||||||||||| |||| || |||||| ||
   247502 TACAAGAATTTGAAGGAGGAAGAAGCACACGGTTTTCTGGAAGGATATTTCCGCCTATGGGGAC 247443

Qy   866 GCATTCCTCCTCCTGCCAGCATTTCCTATCGGGCCAACACAGGCATCTCTTTTAAAGTC 925
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
   247442 GCATTCCTCCTTTGCCAGCCTTTCCTATCGGGTCAACACCGGCATCTCTTTTAAAGTC 247383

Qy   926 TACCAAACACTCAAAGAGTCAAAGATGAGGCAAAAGGTTCTCTTCTTCCATCCCAGGTAC 985

```
Qy      3085 CTGCTAATGAACCTAATTAGCTTAAATTATCTCCTAGCAACATTGGTCACGTTTCAATC 3144
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    245183 CTGCTAATGGGTTAATTAGCTTAAATTATCTCCTAGCAACATTGGTCATGTTTCAATC 245124

Qy      3145 ATGCTATTAGCAAAAAA 316
             |||||||||||||||
Db    245123 CCACTATTTGCAAAGAAA 245 06

RESULT 8
AJ699423
LOCUS        AJ699423                                              197 bp    mRNA        linear    ROD 11-JUL-2005
DEFINITION   Rattus norvegicus mRNA for alpha-2,8-sialyltransferase (SIAT8F
             gene).
ACCESSION    AJ699423
VERSION      AJ699423.1    GI:47057304
KEYWORDS     alpha-2,8-sialyltransferase; ST8Sia-VI.
SOURCE       Rattus norvegicus (Norway rat)
  ORGANISM   Rattus norvegicus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
             Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE    1
  AUTHORS    Harduin-Lepers,A., Mollicone,R., Delannoy,P. and Oriol,R.
  TITLE      The animal sialyltransferases and sialyltransferase-related genes:
             a phylogenetic approach
  JOURNAL    Glycobiology 15 (8), 805 817 (2005)
   PUBMED    15843597
REFERENCE    2  (bases 1 to 1197)
  AUTHORS    Oriol,R.
  TITLE      Direct Submission
  JOURNAL    Submitted (04-MAY-2004) Oriol R., U504, Inserm, 16 Av. Paul
             Vaillant-Couturier, 9480, FRANCE
FEATURES             Location/Qualifiers
     source          1..1197
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon 10116"
                     /clone="ST8Sia-I"
     gene            1..1197
                     /gene="SIAT8F"
     CDS             1..1197
                     /gene="SIAT8F"
                     /function="adds sialic acid to another sialic acid"
                     /codon_start=1
                     /product="alpha-2,8-sialyltransferase"
                     /protein_id="CAG27885.1"
                     /db_xref="GI:47057305"
                     /db_xref="GOA:Q6ZXC7"
                     /db_xref="InterPro:IPR001675"
                     /db_xref="InterPro:IPR012163"
                     /db_xref="UniProtKB/TrEMBL:Q6ZXC7"
                     /translation="MPGGTLFALVGSLMLLLLRLLFWCPADAPARSRMLGESREET
                     GGTSAALKTLWSPTTVPRTRNSTYLDEKTPETAEKCKGLKYSLDSLANKTRYSEDD
                     YLQIVTNIQRCPWKRLEEYENFRAKLASCCDAIQDFVVSQNNTPVGSNMTYEVESKK
                     QIPIRENIFHMPPVSPFVDYPYNQCAVVGNGGILNKSLCGAEIDKSDFVFRCNLPPI
                     TGSASQDVGSKTNLVVNPSIITLKYKNLKEKGARPLEDISAYGDAFLLPAFSYRVN
                     TGISFKVYQTLKESKRQKVLFFHPRYLRHLALFWRTKGVTAYRLSTGLMIASIAVEL
                     CENVTLYGFWPFSKTEBIPLSHHYDNKLPKHGFHOMPKEYSQMLQLHMKGILKLQF
                     SKCBAA"
ORIGIN
Query Match             32.5%;  Score 1028;  DB 6;  Length 1197;
Best Local Similarity   91.2%;  Prd. No. 0;
Matches 1091;  Conservative 0;  Mismatches 105;  Indels 0;  Gaps 0;

Qy        77 ATGAGATCGGGGGGCACGCTGTCGCCCTCATAGGCAGCCAGCTGATGCTGCTCCTCCTG 136
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 ATGCGACCCGGGGGTACGGCTGTCGCCCTGGTAGGCAGCCAGCCTGATGCTGCTCCTCCTG 60

Qy       137 CGTATGCTCTGGTGCCCCGCCCGCCGGCGCCCGCCTCCAGGCTGTGTTGATGGAGGAAGC 196
```

Db   246305 AGTAGTTCCAGTATCTGCTACTTCTCACAGGTTATCATATCCAGGAAGTACCGGCCGA 246246
Qy   2034 --TTTTATTTAAAATAGTGAACATCTACTTGATATCAGCCCGAGGAGGACCATCCTCCATTG 2091
Db   246245 CTTTTTTTTTTTAATAGTAAACATCTATTTGGATATCAGACCTGGGGACTG------TTG 246193
Qy   2092 GAGAATATGAAGAGATATTGTCACTGGCAGAAAAGCAGGTGTGTGCCATTAATTGATAGAT 2151
Db   246192 GAGGATATGAAGAGAGATTGTCACTGGCAGAAGAGCGAGTGTGTGTACCATTAATTGATAAGAC 246133
Qy   2152 AC---CACAAGCATCATCATGCCAGTTATGAACACAGTGCTGAAAAGGATCATAGAACAGGG 2208
Db   246132 ACGGTAGCAGGGCCCATCGTGCAGAGTTATGAGCACAGTGCTGAAAGGATCATAGGCAGGA 246073
Qy   2209 GTGGTTAAATCTGATCCCAGTAGTAATAAACTTCAGTGTACCTATTTCAGGGAAGAGTTAA 2268
Db   246072 ACAGTTAAAATCTGATTCCAGTAGAATCAATGCCAATATGTCTATTTCAGGGAAGAGTTGA 246013
Qy   2269 TTTCACAATTAAAACTAGTAAATGAACCAATTCTTAGCACATTAAGTGGATTCTCGAGTA 2328
Db   246012 TTTCACAATTAAAACTAGAAAACGAACCAATTCTTA-GCAAGTTAAGTGGATTCGGGGTA 245954
Qy   2329 AAAGAAAG-----GGAACAGCAGGAGAAAGCTGTTCGCTTGGTTCGATTACCCAAATGA 2383
Db   245953 AAAAAAAAAAAATGAAACAAGAGGAGAAGGCCGTTTCCTCGACTATGATTTTCCAAATGA 245894
Qy   2384 GCATGCTGGAAGGAGGTGTGAGGCTACGCTAAAACCTCTGCGTAGGGAGAGAGTACAGT 2443
Db   245893 GTATGCTGATAGTGACTGTGAGGTTGCCTGACACACTCGTGTTAGGAGAGAGTACAGG 245834
Qy   2444 GCATGAGTGTGGCGGC-TTTTGTCCACACTCGT---GAAGGGTGAGTAATTCAGAGCCAA 2499
Db   245833 GCATGGGGGTGGTGAGCTTTTTGTCCACACTGGTGAAGAGTGTGAGTAATTCAGAGCCAA 245774
Qy   2500 TCACATCACA---------AGGATGACACACCTAACTCACTTCAGGGGGAGATGA 2549
Db   245773 TCACTTCACAAGGAAGGCTTAGGATAGCGCCTGACACCATTACTTCCAGGGAAAGATGA 245714
Qy   2550 ATGTCTTCATGAGAAATTACACACTCATAGCTAAGAGCATCAGTTTGAGTAAAATTTGAGTA 2609
Db   245713 ACGTTTTCATGAAGAAAATTGCACTCATTGGCTAAGCATCAGTCTGAGTAAAATTTGAATA 245654
Qy   2610 GATGTTAAATAGAACATTTTATACCTCTTACTAAGTGTCCCACCGACACCTTTAATGTA 2669
Db   245653 GATGTTAAATAGAGCAATTTACACCCCTTAC----GTCCCAGGGATACCCCCTTAATGTA 245598
Qy   2670 AGCACATTTATTTATTAAGTTACTTGACATTAAAATGCTATGTCTGTATATTC-TGTTCA 2728
Db   245597 AGCACATTTCTTAGGTTATTTGACT--TATTACATATTTATGTGTGTACATTCTTGTTCA 245540
Qy   2729 TCCATCGATTTTCCCAAAAGTAAGAGCATAGGAGATGAGCCCTACATGCCAAGAAAACT 2788
Db   245539 TCCACGGATGTTCCCAAGAACTAGGAGTCTAGGAGATGCGGTCCACTTGCCAAGAAAACT 245480
Qy   2789 ATAAATTTACTCTTTAATTCTTACTTCTTACTTGAGCCAGCTTGTTGTTTATCAAGTTGTCTTTTTG 2848
Db   245479 ATAAATTTACTCTTTACCTCTCACCTCACCTTGAGCCAGCCGCATTATTCA-CTGATACTTTCCTTG 245421
Qy   2849 AAGAGACAGCACCCTGTGAATTCTTCATTCTGATACAGTGTCACCTTGTATTTAACATTT 2908
Db   245420 AAGGAAGATTACCCCTGTGAATTCTTCATTCTGATACAGTGTCACCTGTATTGTGACATTT 245361
Qy   2909 GTAATGTTGTTTCAAGTTACATCTCTTTTTCATTCTTTTTATAGCAAAATCAAACGTATTAGC 2968
Db   245360 GTAA---CGTGTCAAGTTTGTGCTTCTCTTTCGTTCTTTTTTATAGCCAATCAAACTTGTTAGC 245304
Qy   2969 TTCAGAAATTTATCAGAAGTTCATATATAATATTTTGCAAAGGGTAAAAGGCTTTTTTG 3028
Db   245303 TTTAGAAATATATCGAAGTTCATATATAAATATTTTGCAAAGGGTAAAAAGGCTTTTTTG 245244
Qy   3029 TTAAATAAAATAAA---ATTTATTATTTCTTCTGATGAATAGAGGCTCTTTATGCTG 3084
Db   245243 TTAAATAAAATCAAATGTATTTATTACCTTCTCTGGTTGAATAGAGAGTTCTTTTATGCTG 245184

Db   1141 CAGTCCCATATGAAGGATCCTCAAACTGCAGTTCAGCAAATGTGAAGCCGCTTA 1196

RESULT 9
AJ621583
LOCUS   AJ621583   1234 bp   mRNA   linear   PRI 11-JUL-2005
DEFINITION   Homo sapiens partial mRNA for alpha2,8-sialyltransferase (ST8SIA VI
             gene).
ACCESSION   AJ621583
VERSION   AJ621583.1 GI:46517895
KEYWORDS   alpha2,8-sialyltransferase; ST8SIA VI gene.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
             Hominidae; Homo.
REFERENCE   1
  AUTHORS   Harduin-Lepers,A., Mollicone,R., Delannoy,P. and Oriol,R.
  TITLE   The animal sialyltransferases and sialyltransferase-related genes:
             a phylogenetic approach
  JOURNAL   Glycobiology 15 (8), 805-817 (2005)
   PUBMED   15843597
REFERENCE   2 (bases 1 to 1234)
  AUTHORS   Harduin-Lepers,A.
  TITLE   Direct Submission
  JOURNAL   Submitted (21-JAN-2004) Harduin-Lepers A., UGSF, Universite de
             Lille 1, UMR CNRS USTL n8576, Villeneuve d'Ascq, 59655, FRANCE
FEATURES   Location/Qualifiers
     source   1..1234
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /map="10p12.31"
                     /cell_line="MCF7"
                     /cell_type="adenocarcinoma mammary cell"
     gene   <1..>1234
                     /gene="ST8SIA VI"
     CDS   <1..>1194
                     /gene="ST8SIA VI"
                     /function="sialic acid transfer"
                     /experiment="experimental evidence, no additional details
                     recorded"
                     /codon_start=1
                     /product="alpha2,8-sialyltransferase"
                     /protein_id="CAF21722.1"
                     /db_xref="GI:46517896"
                     /db_xref="GOA:P61647"
                     /db_xref="UniProtKB/Swiss-Prot:P61647"
                     /translation="MPGALLALASLLLLLLRLLWCPADAPGRARILVEESREAT
                     HGTPAALRTLRSPATVPRATNSTYLNEKSLQLTEKCKNLQYGIESFSNKTKGYSEND
                     YLQIITDIQSCPWKRAEEYANFRAKLASCCDAVQNFVVSQNNTPVGTNMSYEVESKK
                     EIPIKKNIFHMFPVSPFVDYPYNQCAVVGNGGILNKSLCGTEIDKSDFVFRCNLPPT
                     TGDVSKDVGSKTNLVINPSIITLKYGNLKEKKALFLEDIATYGDAFFLLPAFSFRAN
                     TGTSFKVYYTLEESKRQKVLFFHPKYLKDLALFWRTKGVTAYRLSTGLMITSVAVEL
                     CKNVKLYGFWPFSKTEDIPVSHHYYDNKLPKHGFHQMPKEYSQIIQLHMKGILKLQF
                     SKCEVA"
ORIGIN

Query Match   27.2%; Score 861.8; DB 5; Length 1234;
Best Local Similarity   82.0%; Prd. No. 3.9e-271;
Matches 1006; Conservative   0; Mismatches 217; Indels   4; Gaps   1;

Qy   77 ATGAGATGGGGGCACGCTGT-CGCCCTCATAGGCAGCCTGATGCTGCTGCTCCTG 136
         |||| ||| |||| || | |||||||||| |||||| |||||| |||||| ||
Db   1 ATGCGACGGCGGGGGCGCACTGCCGCCCTGCTCGCTGCCGCTGCTGCTGCTGCTG 60

Qy   137 CGTATGCTCTGTGTGCCCAGCCGA-CGGGCCTGCCAGCCGCTCCAGGCTGTTGAT... 196
         ||||| ||| |||| ||| || | ||||| |||||| || ||||| ||| ||
Db   61 CGCCTGCTCTGGTGCCCGGCAGC-CGCGGCCGCGGGCGCGCCAGGATTCTGGTGG... 120

Qy   197 AGAGAGGACACCAGTGGTACCTCGGAGGCCGACACCCG 256
         ||||| |||| || || || || |||| || ||||

RESULT 10
AJ697663
LOCUS   AJ697663   197 bp   mRNA   linear   PRI 11-JUL-2005
DEFINITION Pan troglodytes mRNA for alpha-2,8-sialyltransferase 8F (siat8F
            gene).
ACCESSION   AJ697663
VERSION   AJ697663.1 GI:46517916
KEYWORDS   alpha-2,8-sialyltransferase 8E; siat8E gene.
SOURCE   Pan troglodytes (chimpanzee)
  ORGANISM Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1
  AUTHORS   Harduin-Lepers,A., Mollicone,R., Delannoy,P. and Oriol,R.
  TITLE   The animal sialyltransferases and sialyltransferase-related genes:
            a phylogenetic approach
  JOURNAL   Glycobiology 15 (8), 805-817 (2005)
   PUBMED   15843597
REFERENCE   2 (bases 1 to 1197)
  AUTHORS   Oriol,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (20-APR-2004) Oriol R., U504, Inserm, 16 Av. Paul
            Vaillant-Couturier, 94807 FRANCE
FEATURES   Location/Qualifiers
     source   1..1197
                     /organism="Pan troglodytes"
                     /mol_type="mRNA"
                     /db_xref="taxon:9598"
     gene   1..1197
                     /gene="siat8F"
     CDS   1..1197
                     /gene="siat8F"
                     /function="adds sialic acid to another sialic acid"
                     /codon_start=1
                     /product="alpha-2,8-sialyltransferase 8F"
                     /protein_id="CAG26901.1"
                     /db_xref="GI:46517917"
                     /db_xref="GOA:P61648"
                     /db_xref="UniProtKB/Swiss-Prot:P61648"
                     /translation="MPGGALLALASLLLLLLRLLWCPADAPGRARILVEESREAT
                     HGTPAALRTLRSPATVPRATNSTYLNEKSLHLTEKCKNLQYGIESFSNKTKGYSEND
                     YLQIITDIQSCPWKRAEBYANFRAKLASCCDAVQNPVVSQNNTPVGTNMSYEVESKK
                     EIPIKNIFHMFPVSPFVDYPYNQCAVGNGGILNKSLCGTBIDKSDFVFRCNLPPT
                     TGDVSKDVGSKTNLVINPSIITLKYGNLKEKKALFLEDIATYGEAFFLPAFSFRAN
                     TGTSFKVYYTLBESKRQKVLPFPHPYKLKDLLAFLWRIKGVTAYRLSTGLMITSVAVEL
                     CKNVKLYGFWFPSKTEDIPVSHHYYDNKLPKGFHQMPKEYSQILQLHMKGILKLQF
                     SKCEVA"

ORIGIN
Query Match   27.1%; Score 857.8; DB 5; Length 1197;
Best Local Similarity   82.3%; Pr.d. No. 8e-270;
Matches 985; Conservative   0; Mismatches 212; Indels   0; Gaps   0;

Qy   77 ATGAGATCGGGGGCCACGCGT CGCCCCTCATAGGCAGCGCTGATGCTGCTCCTCCTG 136
              ||||| |||||||||||| ||| |||||||||||||||| |||||||||||||||||||||
Db   1 ATGCGGCCCGGGGGCGCACTGC CGCCCCTGCTCGCCAGCGCTGCTGCTGCTGCTGCTG 60

Qy   137 CGTATGCTCTGGTGCCCAGCCG CGGCGCCGGGCCGGGCGCGCAGGATTCTGGTGGAGGAAGC 196
              ||| ||||||||||||||||||| ||||||||| ||||||||||||||| |||||||||||
Db   61 CGCCTGCTCTGGTGCCCGGCCGAC CGCGGCGCCGGCCCGGGCGCGCAGGATTCTGGTGGAGGAAGC 120

Qy   197 AGAGAGGAGCACCAGTGGTACCT AGTCGCACTGAAGACACTCTGGAGCCCGACAACCCCG 256
              |||||||||||| |||||||||| ||||||||||||||||||||| ||||||||||||||
Db   121 AGGGAGGGCCACCCACCGGCACCC CGCAGCGCCTGAGGACGCGCTCCCGGAGCCCGGACCGCG 180

Qy   257 GTACCACGCACCAGGAACAGCA TATATCTGGATGAGAGAACACCCAATAACAGAGAAA 316
              ||||||||||||||||| ||||| ||||||||||||||| |||||||||||||||||||
Db   181 GTACCGGCGCGCCACTAACAGCA ATATCTGAATCGAATGAGAGTCACTCCACCTGACGGAGAAA 240

Qy   317 TGCAAAGATCTGCAATATAGCT GAACTCTTTATCTAACAAAACGAGACGTACTCCTGAG 376

ORGANISM   Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1
AUTHORS   Harduin-Lepers,A., Mollicone,R., Delannoy,P. and Oriol,R.
TITLE   The animal sialyltransferases and sialyltransferase-related genes:
            a phylogenetic approach
JOURNAL   Glycobiology 15 (8), 805-817 (2005)
PUBMED   15843597
REFERENCE   2 (bases 1 to 1197)
AUTHORS   Harduin-Lepers,A.
TITLE   Direct Submission
JOURNAL   Submitted (13-DEC-2004) Harduin-Lepers A., Universite de Lille 1,
            UGSF, UMR CNRS no8576, Villeneuve d'Ascq, 59655, FRANCE
FEATURES   Location/Qualifiers
     source   1..1197
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
     gene   <1..>1197
                     /gene="siat8F"
     CDS   <1..>1194
                     /gene="siat8F"
                     /note="ST8Sia V"
                     /codon_start=1
                     /product="alpha-2,8-sialyltransferase"
                     /protein_id="CAA30813.1"
                     /db_xref="GI:56552215"
                     /db_xref="GOA:Q5NDG0"
                     /db_xref="InterPro:IPR001675"
                     /db_xref="InterPro:IPR012163"
                     /db_xref="UniProtKB/TrEMBL:Q5NDG0"
                     /translation="MPGGALPALLASLLLLLLRLLWCWTDAPARSRFLVESREAT
                     HSSPAAPRKLRSSATPPRTGNSTYLNEKSLHLMERCQNLQAGLETLSNRTKGYSEDE
                     YLQIITNIQSCPWKRVEEYENFRAKLASCCDAVQNFIVSQNNTPIGTNMTYEVESKS
                     KIQIKENIFDMLPVVPFVGYSFNQCAVVGNGGILNQSLCGABIDKADFVFRCNLPPT
                     TGNVINDVGTKTNLVLNPSIIKLRYGNLKEKKAIPLEDVAAYGDAFVLLPAFSFRAN
                     TAASPKVYYTLKESKRQKVLFFHPKYLKNLALFWRTKGVTEYRLSSGLMITSVAVEL
                     CENVKLYGFWPFSRTENMPVSHHYDNKLPKRGFHEMPKEYSQILQHVKGILKLQF
                     SKCETA"
ORIGIN

Query Match   23.9%; Score 757; DB 14; Length 1197;
Best Local Similarity   77.0%; Prd. No. 1.3e-236;
Matches 922; Conservative 0; Mismatches 275; Indels 0; Gaps 0;

Qy   77 ATGAGATCGGGGGCCACGCTGT CGCCCTCATAGGCAGCCTGCTGATGCTGCTCCTCTG 136
          ||||||||||||||||||||||| || |||| ||||||||||||||||||||| ||| ||
Db   1 ATGCGGCCGGGGGGCCGCACTGT CGCCCTGCTCGCCAGCCTGCTGCTGCTGTTGCTGCTT 60

Qy   137 CGTATGCTCGGTTGCCCAGCCG CGGGCCGCTCCGCCGCTGTTGATGGAGGGAAGC 196
          |||||||||||||||||||||| ||||| |||||||||||||||||||||||||||
Db   61 CGCCTGCTCTGGTGCTGGACCG CGGCGCCCGATCCAGGTTCTTGGTGGAGAAAGC 120

Qy   197 AGAGAGGACACCAGTGTACCT GAGTGCACTGGAAGACACTCTGGAGCCCGACAACCCG 256
          ||||||||||||||||||||| |||| |||||||||||||||||||||||||||||||
Db   121 AGGGAGGCCACCCACCAGCAGCC CGCTGCTCCAAGGAGAGTCTCCGGAGCTCGGACCCTG 180

Qy   257 GTACCACGGCACCAGGAACAGCA ATATCTGGATGAGAGACAGAGACAGAGAAA 316
          ||||||||||||||||||||||| || |||||||||||||||||||||||||||
Db   181 CCCCCGGCGCACGGGGAACAGCA ATATCTGAATGAGAATGAGAGTCCCTCCACCTGATGGAGCGC 240

Qy   317 TGCAAAGATCTGCAATATAGCT GAACTCTTATCTAACAAAACGAGACGGTACTCTGAG 376
          |||||||||||||||||||||| ||||| ||||||||||||||||||||||||||||||
Db   241 TGCCAAAATCTGCAAGCGGGGCT GGAGACTTTATCTAACAGAACGAACGAAAGGGTATTCAGAG 300

Qy   377 GATGACTACCTCCAGACCATCA AAACATACAGAGATGCCCATGGAACCGGCAAGCAGAA 436
          |||||||||||||||||||||| | |||||||||||||||||||||||||||||||
Db   301 GATGAGTACCTTCAGATTATCA AAATATCCAGAGCTGTCCATGAAACGACAAGTAGAA 360

Qy   437 GAATATGACAAATTTAGACAA ACTGGCTTCCGTTGCGATGCCATTCAAGACTTCGTG 496
          |||||||||||| |||||||| ||||||||||||| |||| |||||

```
FEATURES            PE Corporation (NY) (US)
      source        Location/Qualifiers
                    1..1149
                    /organism="Homo sapiens"
                    /mol_type="unassigned DNA"
                    /db_xref="taxon:9606"
ORIGIN

Query Match              17.7%;  Score 559.8; DB 2;  Length 1149;
Best Local Similarity    75.5%;  Prd. No. 1.1e-171;
Matches  754;  Conservative  0;  Mismatches 152;  Indels  93;  Gaps  1;

Qy     275 AGCACATATCTGATGAGAGA AAACCCAAATAACAGAGAAATGCAAAGATCTGCAATAT 334
               ||||||||||||||||||||    |||||||| ||||||||||||||||||||||||||
Db     244 AGCACATATCTGAATGAGAAGT GCTCCCACTGACGGAGAAAATGCAAAAATCTGCAATAT 303

Qy     335 AGCTTGAACTCTTTATCTAACA AAACGAGACGGTACTCTGAGGATGACTACCTCCAGACC 394
               ||||||||| | |||||||||    ||||||||||||||||||||||| |||||||||
Db     304 GGCATTGAGTCTTTCTCTAACA AACGAAAGGGTATTCAGAGACGACTACCTTCAGATT 363

Qy     395 ATCACAAACATACAGAGATGCC ATGGAACCGGCAAGCAGCAGAAGCAATATGACAATTTTAGA 454
               |||||||||||||||||||    ||||||||||||||||||||||||||||||||||| |||
Db     364 ATCACAGATATACAGAGTTGTC ATGGAAACGGCAAGCAGCAGAGAATATGCAAATTTTAGA 423

Qy     455 GCAAAACTGGCTTCCTGTTGCG TGCCATTCAAGACTTCGTGGTTTCCCAGAACAACACT 514
               ||||||||||||||||||||    | || ||||||||||| |||| ||| |||||||||
Db     424 GCCAAACTTGCTTCCTCGCGTG TGTCGTTCAAAACTTGTGTTGTTCTCAGAATAACACT 483

Qy     515 CCAGTGGGGACTAAACATGAGCT CGAGGTGGAAGCAAGCAGACATCCCCATTCGAGAG 574
               |||||||||||||||||||||    ||||||||||||||||||||||||||||||||||
Db     484 CCAGTTGGGACTAATATGAGTT CGAGGTGGAAAGCAAAAAAGAAATCCCAATTAAGAAG 543

Qy     575 AACATTTTCCACATGTTCCAG GTCGCAGCCTTTTGTTGTGGACTATCCCCTATAACCAGTGT 634
               |||||||||||||||||||    ||||||||||| ||||||||| |||||||||||||||
Db     544 AACATTTTTCATATGTTTCCAG GTCCCAGCCTTTTGTTGTGGACTACCCTTATAATCAGTGT 603

Qy     635 GCAGTGGTTGGTAATGGGGAA TCTCAACAAGTCTCTCTGCGGAGCAGAAATTGTAAA 694
               ||||||||||||||||||||    |||||||||||||||||||||| ||||| |||||
Db     604 GCAGTGGTTCGGAAATGGGGAA TCTGAATAAGTCTCTCTGTGTGGAACTGAAAATAGATAAA 663

Qy     695 TCTGACTTCGTCTTCAGGTGTA CCTCCCCCAATCACAGGGAGCGGCTAGTAAAGATGTT 754
               ||||||||||||||||||||    |||||||||||||||||||||||||||||||||||
Db     664 TCCGACTTCGTTTTTAG---  --------------------------- 680

Qy     755 GGAAGCAAAACAAATCTTGTGA TGTCAATCCCAGCATTATAACCCTGGAAGTACCAGAAT 814
               || ||| ||| ||||||||||    |||||||||||||||||||||||||||||||||||
Db     681 ------------------  ------ATATGGGAAC 690

Qy     815 TTGAAGGAGAGAAAGCACAGT TTTGGAGGACATCTCCACCTATGGAGAGATGCATTCCTC 874
               |||||||||||||||||||    |||||||||||||||||||||||||||||||||||||
Db     691 TTAAAGGAAAAAAAAAAGCCCTAT TCTGGAGGACAAACTATGCAACCTATGGAGAAATCAGG 750

Qy     875 CTCCTGCCAGCATTTTCCTATC GGCCAACACAGGCATCTCTTTTAAGTCTACCAAACA 934
               |||||||||||||||||||    ||||||||||||||||||||||| ||||||||||||
Db     751 CTTCTGCCAGCATTTTCCTTCA GGCCAACACCGGGTACCTCTTCAAAGTATACTACACG 810

Qy     935 CTCAAAGAGTCAAAAATGAGGC CAAAGGTTCTCTTTCTTCCATCCCAGTGACCTACCTGACAC 994
               |||||||||||||||||||    |||||||||||||||||||||||||||||||||||
Db     811 CTCGAAGAGTCTAAAGCAAGAC AAAGGTTCTATTTTTCCATCCCAAGTGACCTACCTGAAAGAT 870

Qy     995 CTCGCTCCTTTTCTGGAGACTA AGGGGTGACTGCATACCGCTTGTCCACAGGCGTTGATG 1054
               |||||||||||||||||||    ||||||||||||||||||||||| |||||||||||
Db     871 CTGGCCCTTTTCTGGAGACTA AGGTGTGACTGCATACCGCTTGTCCACCGGGCTTGATG 930

Qy    1055 ATTGCAAGTGTCCGTGTGGAAC GTGAAAACGTGAAGCTCTACGGATTCTGGCCTTTC 1114
               || |||||||||||||||||    |||| |||||||||||||||||||||||||||
Db     931 ATCACAAGGTTGCAGTGACATC GTGAAAAATGTGCAGTGAACTGGAATTCTGGCCCTTC 990

Qy    1115 TCTAAGACTATCGAAGACACCC ACTCAGTGACTATCCAGGATAAACATGTTACCTAAG 1174
               |||||| ||||||||||||||    ||||||||||||||||||||||||| ||||||||
Db     991 TCTAAACTGTAGAAGACATAC GTGTCAGCCATCCACTATTATGACAACAAGCTACCTAAA 1050

Qy    1175 CATGGTGTTCCACCAGATGCCTA AGAATACAGCCAAATGCTCCAGTCCCATATGAGAGGA 1234
```

```
Db     361 GAATATGAGAATTTAGAGCAGCAAAACTTCGCTTCCTGTTGTGATGCTGTTCAAAACTTCATT 420
               | ||||  |||||||||||||||||||||||||
Qy     497 GTTCCCAGAACAACAACACTCCCAGTGGGGGGACTAACATGACTGACTGAGGAGGTGAAAGCAAGAAA 556
               |||| || ||||| |||||||||||||||||||||||   ||||||||||||||||||
Db     421 GTTTCTCAAAACAACAACACTCCCAGTTGGGACCAACATGACGTATGAGGTGGAAAGTAAAAGC 480
               ||||||  |||||||| |||||||||  ||||||||
Qy     557 CACATCCCATTCGAGAGAACATTTTCCACATGTTCCAGTGTCGCAGCCTTTTGTTGTGGAC 616
               | |||  |||||||||||||||  |||||||||||||||||
Db     481 AAAATCCAAATTAAAGAAAAACAATTTTGATATGTTGATATGTTGCCAGTGGTGCAGCCTTTTGTGTGGGA 540
               ||   |||||| |  | ||||||||||||||| |||||||
Qy     617 TATCCCTATAACCAGTGTGCAGTGGTTGGTAATGGGGAATTCTCAACAAGTCTCTCTGC 676
               ||  |||    ||||||||||||||||||||||||| || |||||||||||||||||||
Db     541 TACTCTTTTTATCAGTGTGCAGTGGTTGGTTGCAGGGGAATTTTAAATCAGTCTCTCTGT 600
               ||||||  |||||||||||||||||||||||||  ||||||
Db     677 GGAGCAGAAATTGATAAATCTGACTTCGTCTTCAGGTGTAACCTGTAACCTCCCCCCAATCACAGGG 736
               ||||||||||||||||||||  | || |||||||||||
Qy     601 GGAGCTGAAATCGATAAAGCTGACTTCGTTTTTAGGTTAATCTCCCCCCACCACAGGA 660
               ||||||| ||||||| ||||||||||||| | || |||||||
Qy     737 AGCGCTAGTAAAGATGTTGGAGCAAAACAAATCTTGTGACTGTCAATCCCAGCATTATA 796
               |||| |||  ||||||| |||  | |||||   |||||| ||||| |||||||||||||
Db     661 AATGTTACCAATGATGTTGGGACTAAAAACCAACCTGGTGACTCTAAATCCCAGTATTATA 720
               |||   | |||||||||||||||||    |||||||||||||||||||| |||||||
Qy     797 ACCCTGAAGTACCAGAATTGAAGGAGAAGAAAGCACAGTTTTGGAGGACATCTCCACC 856
               |||||| ||  ||||| |||||||||||||||   |||  |||||||||||| ||||||
Db     721 AAGCTAAGATATGGGAAGAACTTAAAGCAAAAGCAATATTTCTGGAGGACGGTTGCAGCC 780
               | ||||||||||||||||||||||| |||||||||||| | ||||||||||||||||
Qy     857 TATGGAGATGCAATCCTCCTGCCAGCACAGGCATTTTCCTATCTGGGCCAACACAGGCATCTCT 916
               |||||||||||||||||  |||  ||||||  |||||||||||||||||||  ||||||||
Db     781 TATGGAGATGCTTTCGTCTTCGTCCTGCTGCGTTTTCCTCAGGGCCAATACCGCGGCCCTCT 840
               |||||  |||||||||||||   |||||||| |||||||||||||||||||||||||
Qy     917 TTTAAGTGTACCAAAACACACAACCTGAAGGCAAAAGTTCTCTTTCTTTCCAT 976
               |||||| ||||| ||||||||||  || ||||||||||||||| | ||| ||||
Db     841 TTTAAAGTGTACTACACTGAAGGAATCAAAGCAACAGCAACAAAAGGTTTTATTTTTCAT 900
               ||   |  |||||| |||||||||  |||||||||||| |||||||||||  |||||
Qy     977 CCCAGGTACCTGAGACACCTCGCTCTTTTCTGGAGAACTGACTGCTCATACCGC 1036
               ||||||||  |||||||| |||||| ||||||||| || || ||||||  |||||
Db     901 CCCAAGTACCTGAAAAAACCTTGCACTCTCTGGAGGACTGAGTGTGACCGGATATCGC 960
               ||||| |||||||||   |||  | || |||||||||  ||||||||||||||||
Qy    1037 TTGTCCACAGGCTTGATGAATTGCAAGTGTCGCTGTGGAACTGTGTGAAAACGTGAAGCTC 1096
               ||||||||||||||||||||||||||   ||||||||||| |||||||||||||
Db     961 CTGTCCTCCGGCTTGATGATCACAAGTGTGGCAGTCGAGCTGTGAAACGTGAAGCTG 1020
               |||||   ||||||||||   |||  |||||||||  |||||||||||||||||||
Qy    1097 TACGGATTCTGGCCTTTCTCTAAGACTATCGAAGACACCACCTCAGTGACCTACTACTAT 1156
               ||||  |||||| | || ||||||||  || ||||||   | || |||||||||||||||
Db    1021 TATGGATTCTGGCCCTTTTCTCTAGAACTGGAGAGAGAACATGCCTGTCAGCCATCACTACTAC 1080
               | |||||||  ||||| ||||||||||||||||| ||||  |||||||||  |||||||
Qy    1157 GATAAACATGTTACCTAAGCATGGTGTTCCACCAGATGCCTCCAAAGAAATACAGCCAAATGCTC 1216
               ||   |||||| | | ||||  ||||||||||| |||||||  ||| ||||||||||||||
Db    1081 GACAATAAGTTACCTAAACGCGGTTTCCATGAGAGATGCCAAAGAGTATAGCCAAAATTCTC 1140
               ||   |||||||||||||||| ||||| ||||||||  ||   |||||  | |||||
Qy    1217 CAGCTCCATATGAGAGGAATCCTCAAACTGCAATTCAGCAAATGTGAAACGGCTAA 1273
               | |||||||||||||||||  |  ||||||||||| |||||||||||||||
Db    1141 CAGCTTCATGTGAAAGGCATCTCCAAGTTACAGTTTAGCAAATGTGAAAACAGCCTGA 1197
```

```
RESULT 12
CQ736679
LOCUS        CQ736679                1149 bp    DNA        linear    PAT 03-FEB-2004
DEFINITION   Sequence 22613 from Patent WO2068579.
ACCESSION    CQ736679
VERSION      CQ736679.1  GI:42332744
KEYWORDS     .
SOURCE       Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
             Hominidae; Homo.
REFERENCE    1
  AUTHORS    Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.
  TITLE      Kits, such as nucleic acid arrays, comprising a majority of
             humanexons or transcripts, for detecting expression and other uses
             thereof
  JOURNAL    Patent: WO 02068579-A 22613 06-SEP-2002;
```

```
Db        1051 CATGGTTTCCATCAGATGCCCAAGAATACAGCCAGATCCTCCAACTTCACATGAAGGA 1110
               ||||||||||  |||||||  ||||||||  |||||||||||||||||| || |||| ||||
Qy        1235 ATCCTCAAACTGCAATTCAGCAAATGTGAAACGGCTTAA 1273
               ||||||||||||  |||||||||||||||||||||  || |||| |||
Db        1111 ATCCTCAAACTGCAATTTAGCAAATGTGAAGTCGCCTAA 1149

RESULT 13
AL158164/c
LOCUS       AL158164         151289 bp    DNA        linear    PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone RP11-414K1 on chromosome 10 Contains
            the 3' end of the gene for the likely ortholog of mouse
            sialyltransferase 8-VI (alpha-2 8-sialytransferase) (ST8SIA-VI), a
            novel gene and the 5' end of a novel gene, complete sequence.
ACCESSION   AL158164
VERSION     AL158164.15  GI:16416157
KEYWORDS    HTG; sialyltransferase; ST8SIA-VI.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 151289)
AUTHORS     Pearce.A.
TITLE       Direct Submission
JOURNAL     Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 24, 2001 this sequence version replaced gi:15028618.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em; EMBL, Sw;, SWISSPROT; Tr;, TREMBL; Wp;, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 10, constructed by the Sanger Centre Chromosome 10
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr10
            RP11-414K1 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            ------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            -------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compresions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
FEATURES             Location/Qualifiers
     source          1..151289
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-414K1"
                     /clone_lib="RPCI-11.2"
     misc_feature    2000
                     /note="Clone right end: RP11-124N14"
     gene            complement(join(<54812..55162,57619..57775,59494..59586,
                     63443..>63538))
                     /gene="RP11-414K1.1"
                     /locus_tag="RP11-414K1.1-002"
     mRNA            complement(join(<54812..55162,57619..57775,59494..59586,
                     63443..>63538))
                     /gene="RP11-414K1.1"
```

```
     CDS             complement(join(55124..55162,57619..57775,59494..59586,
                     63443..>63538))
                     /gene="RP11-414K1.1"
                     /standard_name="OTTHUMP00000019236"
                     /codon_start=2
                     /product="likel  ortholog of mouse sialyltransferase 8-VI
                     (alpha-2, 8-sia ytransferase) (ST8SIA-VI)"
                     /note="match: E Ts: BP678643.1"
                     /protein_id="CA 40659.1"
                     /db_xref="GI:57 08950"
                     /db_xref="GOA:Q JUL6"
                     /db_xref="UniPr  KB/TrEMBL:Q5JUL6"
                     /translation="Y YNQCAVVGNGGILNKSLCGTBIDKSDFVFRCNLPPTGDVSK
                     DVGSKTNLVINPSI TLKYGNLKEKKALFLEDIATYGDAFFLLPAFSFRANTGTSFK
                     VYYTLESKARQKCK WRKTINSIL"
     gene            join(complement AL160289.14:41295..41470),
                     complement(AL16 289.14:40699..40797),
                     complement(1269 0..127049),complement(95943..96029),
                     complement(6783  ..67981),complement(63443..63555),
                     complement(5949  ..59586),complement(56303..57775))
                     /gene="RP11-414  1.1"
                     /locus_tag="RP1  -414K1.1-001"
     mRNA            join(complement AL160289.14:41295..41470),
                     complement(AL16 289.14:40699..40797),
                     complement(1269 0..127049),complement(95943..96029),
                     complement(6783  ..67981),complement(63443..63555),
                     complement(5949  ..59586),complement(56303..57775))
                     /gene="RP11-414  1.1"
                     /locus_tag="RP1  -414K1.1-001"
                     /product="likel  ortholog of mouse sialyltransferase 8-VI
                     (alpha-2, 8-sia ytransferase) (ST8SIA-VI)"
                     /note="match: E Ts: BB627781.1 BB660248.1 BE686184.1
                     BG919404.1 BI10 002.1
                     match: cDNAs: A  059954.1"
     CDS             join(complement AL160289.14:41295..41395),
                     complement(AL16 289.14:40699..40797),
                     complement(1269 0..127049),complement(95943..96029),
                     complement(6783  ..67981),complement(63443..63555),
                     complement(5949  ..59586),complement(57307..57775))
                     /gene="RP11-414  1.1"
                     /locus_tag="RP1  -414K1.1-001"
                     /standard_name= OTTHUMP00000019235"
                     /note="match: p oteins: CAA11685 O15466 O43173 P70126
                     P70554 P79783 P 7877 Q64468 Q64687 Q8JZQ3 Q8K4T1 Q92185
                     Q9CUU6"
                     /codon_start=1
                     /product="likel  ortholog of mouse sialyltransferase 8-VI
                     (alpha-2, 8-sia ytransferase) (ST8SIA-VI)"
                     /protein_id="CA 40658.1"
                     /db_xref="GI:57 08949"
                     /translation="M PGGALLALLASLLLLLLRLLLWCPADAPGRARILVESREAT
                     HGTPAALRTLRSPAT VPRATNSTYLNEKSLQLTEKCKNLQYGIESFSNKTKGYSEND
                     YLQIITDIQSCPWKR AEEYANFRAKLASCCDAVQNFVVSQNNTPVGTNMSYEVESKK
                     EIPIKNIFHMPPVS PFVDYPYNQCAVVGNGGILNKSLCGTBIDKSDFVFRCNLPPT
                     TGDVSKDVGSKTNLV INPSIITLKYGNLKEKKALFLEDIATYGDAFFLLPAFSFRAN
                     TGTSFKVYYTLBESK RQKVLFFHPKYLKDLLAFWRTKGVIAYRLSTGLMITSVAVBL
                     CKNVKLYGFWPFSKT EDIPVSHHYDNKLPFGHFHQMPKEYSQILQLHMKGILKLQF
                     SKCEVA"
     misc_feature    85117
                     /note="Clone le t end: RP11-161K20"
     gene            complement(join 91406..91512,91566..91792,95943..96029,
                     103847..103960 126960..126995))
                     /gene="RP11-414  1.1"
                     /locus_tag="RP1  -414K1.1-003"
     mRNA            complement(join 91406..91512,91566..91792,95943..96029,
                     103847..103960 126960..126995))
                     /gene="RP11-414  1.1"
                     /locus_tag="RP1  -414K1.1-003"
                     /product="likel  ortholog of mouse sialyltransferase 8-VI
```

```
                    (alpha-2, 8-sialytransferase) (ST8SIA-VI)"
                    /note="match: ESTs: AV714822.1"
gene            join(123423...123504,135534...135608,144397...144715)
                    /locus_tag="RP11-414K1.3-001"
gene            join(123423...123504,135534...135608,144492...144715)
                    /locus_tag="RP11-414K1.3-002"
mRNA            join(123423...123504,135534...135608,144397...144715)
                    /locus_tag="RP11-414K1.3-001"
                    /product="novel transcript"
                    /note="match: ESTs: AI953053.1 AW139127.1 AW340196.1
                    BI826833.1 BU537392.1
                    match: cDNAs: BC029655.1"
mRNA            join(123423...123504,135534...135608,144492...144715)
                    /locus_tag="RP11-414K1.3-002"
                    /product="novel transcript"
                    /note="match: ESTs: AA813705.1 AI632435.1 AW449205.1
                    BE045987.1"
gene            join(144469...144673,149779...149932)
                    /locus_tag="RP11-414K1.2-001"
mRNA            join(144469...144673,149779...149932)
                    /locus_tag="RP11-414K1.2-001"
                    /product="putative novel transcript"
                    /note="match: ESTs: AA609183.1"
polyA_signal    144692...144697
                    /locus_tag="RP11-414K1.2-001"
polyA_site      144715
                    /locus_tag="RP11-414K1.2-001"
misc_feature    151289
                    /note="Clone_right_end: RP11-414K1"
ORIGIN

Query Match    14.9%;    Score 470.4;    DB 5;    Length 151289;
Best Local Similarity  64.7%;    Pred. No. 6.1e-142;
Matches 1008;    Conservative    0;    Mismatches 481;    Indels    69;    Gaps    18;

Qy    740  GCTAGTAAAGATGTTGGAAGCAAAACAAATCTTGTGACTGTCAATCCCAGCATTATAACC  799
           |||||||||| ||| || ||||||||| |||||| |||||| |||||| ||||| || ||
Db  57840  GTTATCAAAACTGCTACTTGATTATTCTTGGGATTACTATAATTTTTTTTTTTTTAATTT  57781

Qy    800  CTGAAGTACCAGAATTTGAAGGAGAAAGCACAGTTTTTGGAGGACATCTCCACCTAT  859
           |||  |  ||  | ||  | |||   ||| || || | |||| ||| |||  |||||
Db  57780  GGCAGAGATATGGGAACTTAAAGGAAAAAAAGCCCTATTTTCTGGAGGACCAAACCTAT  57721

Qy    860  GGAGATGCATTCCTCCTGCCAGCATTTTCCATGGGCCAACACAGGCATCTCTTTT  919
           |||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db  57720  GGAGATGCATTTTTCTCTGCGCAGCATTTTCCTTCAGGGCCCAACACGGGTACCTCCTTC  57661

Qy    920  AAAGTCTACCAAACACTCAAAGAGTCAAAAATGAGGCAAAAGTTCTCTTCTTCCATCCC  979
           ||||| || |||| ||||| ||||||  ||| |||||||||| || ||||| |||||||
Db  57660  AAAGTATATACACACGCTCGAAGAGTCTAAAGCAACAAAAGGTTCATTTTTCCATCCC  57601

Qy    980  AGGTACCTGAGACACCTCGCTCCTTCTCTGGAGACTAAAGGGGTGACTGCATACCGCTTG  1039
           |||||  |||||||||| | ||| |||||| |||| |||||||  |||| |||||||||
Db  57600  AAGTAGTACCTCGAAGAGTCTAAAGGTGTGACTGCATACCGCTTG  57541

Qy   1040  TCCACAGGCTTGATGATTGCAAGTGCTGTGGAACTGTGTGAAAACGTGAAGCTCTAC  1099
           ||||||||||||||| ||||||||| ||||| ||||||||||| |||||||||||||
Db  57540  TCCACCGGCTTGATGATCACAAGTGTTGCAGTGCAATGTGTAAAAATGTGAAGCTGTAT  57481

Qy   1100  GGATTCTGGCCTTTCTCTTAAGACTATCGAAGACACCCCACTCAGTCACTACTATGAT  1159
           |||||| || |||||||| |||| ||| |||| |||| ||| || |||| ||||||||
Db  57480  GGATTCTGGCCCTTCTCTAAAACTGTAGAGAGACATACCGTCTCAGCCATCTATTATGAC  57421

Qy   1160  AACATGTTACCTAAGCATGGTTTCCACCAGATGCCTAAGAATACAGCCAAATGCTCCAG  1219
           |||||| |||||||||||| | ||||||||||| | ||| ||||| || ||| |||||
Db  57420  AACAAGCTACCTAAACATGGTTTTCCATCAGATGCCCAAAGAATACAGCCAGTCCTTCCAA  57361

Qy   1220  CTCCATATGAGAGGAATCCTCAAACTGCAATTCAGCAAATGTGAAACGGCTT----AACG  1275
           ||| || |||| || | || ||| ||||| |||| ||| ||||||  |||    ||||
Db  57360  CTTCACATGAGAAGGAATCCTCAAACTGCCAATTTAGCAAATGTGAAGTCGCGCCTAAACAAAG  57301

Qy   1276  TTTCTTAGAAGGAGAATAATTTCAGGA-GGTGGAGTGGATGGTGTCACAGCATCTCCAAAA  1334
           ||||| ||| ||| || ||| |||||   || |||  || |  || || ||| ||||||||
```

Db   57300 TATCTTAAAAATGGGAATAATTTAATATATGCAGTAGGTGATTGATTAACAATGTCTCCAAAC 57241

Qy   1335 AGCCAATAGAGAAGAGAGGCACAGA-----AAAGCATGCATGAATTACAAAGGCGCTCTCCCCACTIGTCTA 1394
              || |||||| || || ||| || | ||| | |||| ||| |||| ||||| | || |||||||| ||||
Db   57240 ACC-------AAAGGAGTGGTTAAAGAGTATTTTGAGATGACGAGCCCCAAAATTTGGTTT 57189

Qy   1395 CACTTTGCAGCTCCAGAGTCACTC-ATTCTCACCT 1453
              ||||||||||| || || | |||| ||||| ||||
Db   57188 GACCAAAGGCACCCGCCCCAC----CTTCCCCAC 57134

Qy   1454 TCAACGTTCTTCTTCCTGAGAAT-GAGACCAAAAACATCAGACTTGGATAAGTAAAATGAGA 1513
              ||||| |||| | | |||||| ||| |||| || || ||| || ||||| | |||||||
Db   57133 TCTTCATTTTTCTTCCTTATAA-ATGGACACCATATCTGACTCATCACAAATTTTATAAT 57074

Qy   1514 TAATTTTTCAAATCATCATAGAGA-TTTGATTGAGCCAGGTCTCCAGAATGCTTCCTTG 1573
              || |||||||| | | | ||||| |||||||||||||||| | || |||||| |||
Db   57073 TTATGAAAATTATTGGCATGGC-TTTGGTGTGAGGTAGGGACTCAAAGATATTTCTTTC 57014

Qy   1574 TTCCTATCATGATAGCCATTC-CACCTTTATCAGAGTGTAATGAAACTGTGCAATTGT 1633
              |||| ||| || | |||||| | |||| ||| ||||||||| || |||||||||||||
Db   57013 TTACTATTTATACCAGCCATCT-CACCTTCACCAGACCGATAATAAAAAGCATGCTTTCTT 56954

Qy   1634 GCCAAAGACCCTTTCTGAAGAG-------CATGGGCGCGAGTTTTTACACA 1687
              ||||||| ||| ||| ||||| ||||||||||| ||||| |||||
Db   56953 G--GAGATCTCTTCTCAAGAGGATAGCTGCATTACCAGTATCCATTGGATTTCTTTCTG 56897

Qy   1688 CAGCTCCTTCCTTTATAAATAAA-CCTTCCCCATTCTCCCCTCAGTAGAGTACAG------ 1741
              ||| |||||||| |||||| | |||||||||||| | ||| |||| |||||
Db   56896 CAGCTCCCATTTTCTGTGTCTCAAAGACTGCATGCAAGA 56837

Qy   1742 ------------AAACAAAT-CCCTGATGATTCAGGAAGAAGAAAGTCTTTTTTACTTA 1788
              ||||||| ||| || |||||| ||| ||| |||||||
Db   56836 AAATTATATACTTAAGCAAAAA-TTCTTCATCATTCGGGAATGAAATTCTTCTCTCGAGTA 56777

Qy   1789 --GCAATGTGCCTGCTTCTGAT-CAGTTCGCTTGTGACATTAAGCTGGGTTGGGGTTTTG 1846
              || | || | ||| ||||| | ||| ||| || || || ||| ||||| |||
Db   56776 TGGCAGTAGCCGCCTTCCTTCGATAAATTCACTTTTAATAGT-AGTGTGATGATATTTTTG 56718

Qy   1847 GTTGAATTGGGGCGTTCTT-CTTCTTTGTCTATATTTCCTTACCTTTATCAGTTT 1906
              |||||||| ||| |||||| |||||||| |||| || || |||| |||||| ||||
Db   56717 GTTGAATTTAAGCCATTTATT-CCTCTTATGTCTCACATTTCATCACTTAATAATTT 56658

Qy   1907 GTATTCGAGCTTCCTGCTTTGG-ATTCTGCAATTCTCTCTCCCACTGACAGGATCAACTC 1966
              |||| ||||||||| | ||||| | ||| ||||| ||| ||||| | ||||||||
Db   56657 -TATTTTAACTTCCTGCCTCAT-CTTACACTTTTTTCTCTCTCCACTCACTTATACGATATACTT 56599

Qy   1967 AATGACATAAAGTAGTTCAAAC-TCCATTGCTTCTCACATGTTTAT---CCATAAAGTT 2023
              ||||||||||||| ||||||| | ||| |||||| ||||||| | | ||||||||
Db   56598 CAAGGACGTGCTGAACAGTTACAAT-CCAGTCTTTCTTACCGTTGTCTGTTGTGGGGTC 56539

Qy   2024 ACTCATCTGATTTTATTTAAAA-TAGTGAACATCTACTTGATATGATCAGAGACCACCA 2081
              || || || ||||| ||||| | ||| || | ||||| |||||| ||||||||||||
Db   56538 ACTCACTCAATTTTATTTTAA-TTACTGAATAACTACTTCATGCCAGGCCCT-GCACCA 56480

Qy   2082 TCCTCCATTGGAGAATATGAAG---TATTGTCACTGGCAGAAAGCAGTGTGTGCCATT 2139
              | ||| |||| | |||| || | ||||| || || || ||| || ||| ||||||||
Db   56479 GGCTTCGTGTGAGAACAGCAACAGAATTATTGCCTTTGTAGGAGTAGGAAACACAGATGCATAT 56420

Qy   2140 AATTGATAA------GATACCA-AGTTATGAACACAGTGCTG 2190
              |||| |||| ||||| | ||| |||| | ||| ||
Db   56419 CATTAATTATGATGGGTGATGA-AAGTTATTACACTCATAAGTGCTA 56360

Qy   2191 AAAGGATCATAGACAGGGGTGG-TAAATCTGATCCCAGTAGAATAAAAACTTCAGTGTAC 2248
              |||||| ||||| |||||| || |||| ||| |||||| | | ||||| |||||||||||
Db   56359 CAAGGAACAGAAGCAGCAGGATAA-CATTTCAGTGGAATCTATTTCAATGTAC 56303

RESULT 14

```
KEYWORDS     alpha-2,8-sialyltransferase; SIAT8F gene.
SOURCE       Gallus gallus (chicken)
  ORGANISM   Gallus gallus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
             Phasianinae; Gallus.
REFERENCE    1
  AUTHORS    Harduin-Lepers,A., Mollicone,R., Delannoy,P. and Oriol,R.
  TITLE      The animal sialyltransferases and sialyltransferase-related genes:
             a phylogenetic approach
  JOURNAL    Glycobiology 15 (8), 805-817 (2005)
  PUBMED     15843597
REFERENCE    2  (bases 1 to 1197)
  AUTHORS    Oriol,R.
  TITLE      Direct Submission
  JOURNAL    Submitted (04-MAY-2004) Oriol R., U504, Inserm, 16 Av. Paul
             Vaillant-Couturier, 94807, FRANCE
FEATURES             Location/Qualifiers
     source          1..1197
                     /organism="Gallus gallus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9031"
                     /clone="ST8Sia-VI"
     gene            1..1197
                     /gene="SIAT8F"
     CDS             1..1197
                     /gene="SIAT8F"
                     /function="adds sialic acid to another sialic acid"
                     /codon_start=1
                     /product="alpha-2,8-sialyltransferase"
                     /protein_id="CAG27886.1"
                     /db_xref="GI:47057307"
                     /db_xref="GOA:Q6ZXC6"
                     /db_xref="InterPro:IPR001675"
                     /db_xref="InterPro:IPR012163"
                     /db_xref="UniProtKB/TrEMBL:Q6ZXC6"
                     /translation="MPAGQRPSGSGRRPVVALAGFDNFCYPVLLARGWEEEAEVATVT
                     PKVVRALRSPLTSLPQTENRTTANKDGIYQPELDTKCKAIQDNILSSSFKKKRYPEDY
                     YLHIVKLQNCTWVRRPEESAKFRSELASCCNAVHNFIASQNNTPLGSNMSYEVDSKK
                     TILITEDIFRMLPVSSPLSVYPFKNCAVVGNGGILKNSSCGAEIDRSDFVFRCNLPPT
                     MGSISKDVGNKTKLVTVNPSIIAQKYNKLNEKKTEFVENVAVYGDAPLLJPAFSFRSN
                     TATSFKVYHTLQEFKATQRAIFFHPTYLKNLAQFWRTKGVKAYRLSSGFMITSAAVEL
                     CENVKLYGFWPFSKSTEKMPISHHYYDNQLPKPGFHAMPKEYNQIIQLHGKGIKKLQF
                     GKCESD"
ORIGIN
Query Match          14.8%;  Score 467.2;  DB 11;  Length 1197;
Best Local Similarity   67.9%;  Pred. No. 3.5e-141;
Matches  652;  Conservative   0;  Mismatches 308;  Indels   0;  Gaps   0;

Qy    314 AAATGCAAAGATCTGCAATATAGCTTGAACTCTTTATCTAACAAAACGAGACGGTACTCT 373
          |||||||  |||||  |  ||||||||||| | |||||||  |||||||  ||| ||||||
Db    238 AAATGTAAGGCAATTCAGGACAACATTTTGTCATCGTCGTTCAAGAAGAAAAGGTATCCA 297

Qy    374 GAGGATGACTACCTCCAGACACCATCACAAACATACAGAGATGCCCATGAACCGGCAAGCA 433
          || ||||| ||||||||||||| |  |  ||  ||  |||||||| ||||  |||||||||
Db    298 GAAGATTATTATCTTCACATAGTCAAAAAAACTGCAGAACTGCACCTGGGTAAGGAGACCG 357

Qy    434 GAAGAATAATGACAATTTAGAGCAAAACTGGCTTCCTGTGCCGATGCCATTCAAGACTTC 493
          ||||||| ||| |  |||  |||  |||| ||| ||| | ||| ||  |||| ||||||||
Db    358 GAAGAATCAGCAGAAATTCAGGTCAGAGTTAGCTTCCTGCTGTAATGCAGTTCAACAATTT 417

Qy    494 GTGGTTTCCAGAACAACACTCCAGTGGGACTAACATGAGCTACGAGGTGGAAAGCAAG 553
          ||||| ||||||| ||||  | ||||||  | |||  |||| ||||  |||||||||||
Db    418 ATTTGCTTCTCAAACAACACACCCCACTGGGAAGTAACATCAGTTATGAAGTGGACAGTAA 477

Qy    554 AAACACATCCCATTCGAGAGAACATTTCCACATGTTCCAGTGTCGCAGCCTTTTGTG 613
          ||||| |||| |||  ||||| |||| |||| |||||||| |  ||  |||| ||||
Db    478 AAGACCATCCATTACAGAAGAAGACATTTTAGAATGCTGCCTGTCTGTCCTCCTTTTCA 537

Qy    614 GACTATCCCTATAACCAGTGTGCAGTGGTTGGTAATGGGAATTCCAACAGTCTCTC 673
          |||||| |  | |||| |||||||||||||||||| ||||  |||| ||||||||
Db    538 GTTTATCCCTTCAAGAACTGTGCAGTGGTTGGTAATGGAGGAATTCTGAAAAATTCCAGC 597
```

```
source          1..1520
                /organism="Gallus gallus"
                /mol_type="mRNA"
                /strain="White Leghorn, Hisex"
                /db_xref="taxon:9031"
                /clone="CHEST321n14"
                /clone_lib="CSEOCHN24"
                /dev_stage="stage 22"
ORIGIN

Query Match      14.4%;    Score 456.8;    DB 11;    Length 1520;
Best Local Similarity   67.9%;   Pred. No. 9.8e-138;
Matches  652;  Conservative   0;  Mismatches  307;   Indels   1;   Gaps   1;

Qy    314 AAATGCAAAGATCTGCAATATAGCTTGAACTCTTTATCTAACAAAACGAGACGGTACTCT 373
          |||| ||| | ||||  |||||| ||| |||  |||||||| || ||||| ||| ||||
Db    229 AAATGTAAGGCAATTCAGACACAACATTTTGTCATCGTCGTTCAAGAAGAAAAGGTATCCA 288

Qy    374 GAGGATGACTACCTCCAGACCATCACAAACATACAGAGATGCCCATGGAACCGGCAAGCA 433
          |  ||| |  ||| ||| |||| ||  |||| ||| ||||| ||||| |||  |||| |
Db    289 GAAGATTATTATCTTCACATAGTCAAAAAACTGCAGAACTGCACCTGGGTAGGGAGACCG 348

Qy    434 GAAGAATATGACAATTTAGAGCAAAACTGGCTTCCTGTTGGGATGCCATTCAAGACTTC 493
          |||||||  ||| |   ||||  |||||   |||| |||| |  |    | ||||| |
Db    349 GAAGAATCAGCAGAAATTCAG-TCAGAGTTAGCTTCCTGCTGTAATGCAGTTCACAATTTT 407

Qy    494 GTGGTTTCCCAGAACAACACTCCAGTGGGACTAACATGAGCTACGAGGTGAAAGCAAG 553
          |  |  |||||  |||| ||  ||||| ||| || ||||| |||| ||||| ||||
Db    408 ATTGCTTCTCAAAACAACACCCCACTGGGAAGTAACATGAGTTATGAAGTGGACAGTAAA 467

Qy    554 AAACACATCCCCATTCGAGAGAACATTTTCCACATGTTTCCAGTGTCGCAGCCTTTGTG 613
          |||||| |||||| | ||||  |||||  |||| ||  | || |||  |||| | |||
Db    468 AAGACCATCCCATTACAGAAGCATTTTTAGAATGCTGCCTGTCTCCTCCTCTTTCA 527

Qy    614 GACTATCCTATAACCAGTGTGCAGTGGTGGTAATGGGGAATTCTCAACAAGTCTCTC 673
          | |  ||| ||  | ||| |||  |  ||| |||||  |||||| |||||  ||||
Db    528 GTTTATCCCATCAAGAACTGTGCAGTCGTGGTTGGAAATGGAGCCATTCTGAAAAATTCCAGC 587

Qy    674 TGCGGAGCAGAATTGATAAATCTGACTTCGTCTTCCAGGTGGTAACCTCCCCCAATCACA 733
          |||  || |  |||| ||||| |||| ||  || |||| |||||  |||| |||| ||
Db    588 TGTGGAGCTGAAATGCATCGTTTGTTTCGGTGTAACCTCCTCCAACAATG 647

Qy    734 GGGAGCGCTAGTAAAGATGTTGGAAGCAAACAAAATCTGTGACTGTCAATCCCAGCATT 793
          |||||  || ||| |||||  ||| |||| ||||||  |||| ||  |||| |||||
Db    648 GGAAGCATTAGCAAAGATGTCGGCAATAAACAAACCTTGTGACGGTTAATCCAAGTATC 707

Qy    794 ATAACCCTGAAGCACCAGAATTGAAGGAGAGAAAGCACAGTTTTGGAGGACATCTCC 853
          || | |||  |||| |  |  || ||| |||||| || |||||| |||| | |||| |
Db    708 ATAGCTCAGAAATACAACAACTAAATGAAAGAAGACAGAATTTGTGGGAAACGTTGCA 767

Qy    854 ACCTATGGAGATGCATTCCTCCTCCGCCAGCATTTTCCTATCGGGCCAACACAGGCATC 913
          |||| ||| | ||| |||  || |  || ||||||   ||  ||||||  ||| ||  |
Db    768 GTCTATGGAGATGCTTTTCTTCCTTTATGCCGGCATTTTCTTCAGAAGCAACAACGGGCTACT 827

Qy    914 TCTTTTAAAGTCACCAAACACTCAAAGAGTCAAAAATGAGGCAAAAGGTTCTTCTTC 973
          |||||||| | |   ||||| |||  | ||  ||| | || ||| || || ||| |
Db    828 TCTTTTAAAGTCATCACACTTGCAGGAGTTCAAAGCAACCCCAAAGGGCAATATTTTTT 887

Qy    974 CATCCCAGGTACCTGAAGACACCTCGCTCTTTCTGGAGACTAAAGGGGTGACTGCATAC 1033
          |||||||| | ||   |   || || | ||  ||||| ||| ||||| || || ||||
Db    888 CACCCCACCTACTTGAAAATCTGAAAAATCTTGCCCAGTTTTGGAGACTAAGGGTGTAAAGCCTAC 947

Qy   1034 CGCTTGTCCCACAGGCTTGGCCTTTCTGATGATTGCAAGTGTCGCTGTGAAAAACGTGAAG 1093
          ||  ||||  ||| | || ||||| || | || ||||  || ||  |||| | ||| |||
Db    948 CGGTTATCATCTGTTTTATGATCACTAGTGCTGTCAGTCGTGTTGAGCGTGTGTGAGAATGTGAAG 1007

Qy   1094 CTCTACGGATTCTGGCCTTTCTCTTAAGACTATCGAAGACACCCCACTCAGTCACCACTAC 1153
          | ||||| ||||| ||||| || |||| || |  ||| ||| ||  ||||| ||||| ||
Db   1008 CTTTACGGCTTCTGGCCTTTCTCAAAATCTCAAATCAGAGATGCCACCACCTAT 1067

Qy   1154 TATGATAACATGTTACCTAAGCATGGTTTCCACCAGATGCCTAAAGAATACAGCCAAATG 1213
          ||||| |  || ||||||   || || |||  ||| ||  | |  | | ||| ||| |
Db   1068 TACGACAACCAGTCGCCCAAACCAGGTTTCCAATGCAATGCCCAAGGAATACAACCAGATC 1127

Qy   1214 CTTCCAGCTCCATATGAGAGGAACCTCAAACTGCAATTCAGCAAAGTGAAAACGGCTTAA 1273
          ||| ||||||| ||||| | | |   ||||| || |||||  ||| |||||| |   |||
Db   1128 CTTCAGCTTCATGGCAAGGCATTTGAAGTTGCAGTTTGGTAAATGTGAATGTGAATCAGACTAA 1187
```

Search completed: May 31, 2006, 15:4:56
Job time : 17619 secs

```
                    GenCore version 5.1.8
           Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on:        May 31, 2006, 10:49:29 ; Search time 14318 Seconds
                                        (without alignments)
                              12364.898 Million cell updates/sec

Title:         US-10-501-930-2
Perfect score: 3166
Sequence:      1 cggagcggcgagtcggtgcc.........gctattagcaaaaaaaaaaa 3166

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched:      48236798 seqs, 2795965780 residues

Total number of hits satisfying chosen parameters:    96473596

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     EST:*
               1: gb_est1:*
               2: gb_est3:*
               3: gb_est4:*
               4: gb_est5:*
               5: gb_est6:*
               6: gb_htc:*
               7: gb_est7:*
               8: gb_est8:*
               9: gb_est9:*
              10: gb_gss1:*
              11: gb_gss2:*
              12: gb_gss3:*
              13: gb_gss3:*
              14: gb_gss4:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

```
                %
Result          Query
No.    Score    Match  Length DB  ID          Description
------ -------- -----  ------ --- ----------   -----------
    1   3068.6   96.9   3222   6  AK172400     AK172400 Mus muscu
    2   3026.2   95.6   3695   6  AK156557     AK156557 Mus muscu
    3   3010     95.1   3292   6  AK085105     AK085105 Mus muscu
    4    818     25.8    820   2  CX565877     CX565877 UI-M-IB0-
    5    760.4   24.0    812   2  BG919404     BG919404 602818295
    6    748.4   23.6    780   2  BI695458     BI695458 603346462
    7    718     22.7    753   2  BI653314     BI653314 603301030
    8    666.8   21.1    670   7  BB486599     BB486599 BB486599
  c 9    631     19.9    640  11  AZ315725     AZ315725 1M0032M24
   10    624.8   19.7    628   7  BB112539     BB112539 BB112539
   11    604     19.1    625   7  BB627781     BB627781 BB627781
  c 12   596.4   18.8    631   7  BE633149     BE633149 uv73e09.y
   13    575.4   18.2    663   4  BY752002     BY752002 BY752002
   14    573     18.1    852   2  BI653246     BI653246 603300931
   15    570     18.0    771   2  BI100002     BI100002 602884742
   16    482     15.2    517   2  BE686184     BE686184 uv73e09.x
   17    468     14.8    936   4  CA976319     CA976319 AGENCOURT
   18    458     14.5    472   7  BB678823     BB678823 BB678823
  c 19   456     14.4    463   7  BE553728     BE553728 ur41e03.x
```

```
   20   453.2  14.3  466  7   BB832306  BB832306
   21   449.6  14.2  483  7   BB829625  BB829625
   22   443.8  14.0  464  4   BY560796  BY560796
   23   443.4  14.0  462  4   BY568870  BY568870
   24   433.2  13.7  470  7   BB823042  BB823042
   25   431.6  13.6  446  4   BY564759  BY564759
   26   424.4  13.4  446  4   BB22186   BB22186
   27   423.6  13.4  439  4   BY568373  BY568373
   28   420.8  13.3  444  7   BB34183   BB34183
   29   420    13.3  483  14  AY404993  AY404993 Mus muscu
   30   416    13.1  482  4   BY562845  BY562845
   31   411.8  13.0  467  4   BY570610  BY570610
   32   402.8  12.7  457  4   BY573730  BY573730
   33   402.6  12.7  477  7   BB829600  BB829600
   34   398    12.6  398  4   BJ368484  BJ368484
   35   390.8  12.3  410  4   BY379395  BY379395
   36   383.4  12.1  439  7   BB833290  BB833290
   37   381    12.0  381  2   BF730564  BF730564 mab72f04.
   38   377.2  11.9  801  3   BU447532  BU447532 603213950
   39   375.2  11.9  388  7   BB836534  BB836534
   40   368.8  11.6  412  4   BY379548  BY379548
   41   358.4  11.3  361  4   BY028633  BY028633
   42   357    11.3  413  4   BY370069  BY370069
   43   349.4  11.0  367  4   BY020379  BY020379
   44   345.2  10.9  383  4   BY024148  BY024148
   45   340.2  10.7  458  7   BB744521  BB744521
```

ALIGNMENTS

```
RESULT 1
AK172400
LOCUS       AK172400                3222 bp    mRNA    linear   HTC 21-SEP-2005
DEFINITION  Mus musculus activated spleen cDNA, RIKEN full-length enriched
            library, clone:F830206C11 product:sialyltransferase 8 (alpha-2,
            8-sialyltransferase) F, full insert sequence.
ACCESSION   AK172400
VERSION     AK172400.1  GI:74186460
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-4  (1999)
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Tagawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ogawa,Y., Tanaka,K., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system-384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
```

AUTHORS   Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
           Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
           Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
           Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
           Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
           Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
           Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
           Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
           Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
           Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
           Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
           Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
           Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
           Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
           Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
           and Hayashizaki,Y.
CONSRTM   RIKEN Genome Exploration Research Group Phase II Team and the
           FANTOM Consortium
TITLE   Functional annotation of a full-length mouse cDNA collection
JOURNAL   Nature 409 (6821), 685-690 (2001)
PUBMED   11217851
REFERENCE 5
AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
           Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
           Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
           Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
           Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
           Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
           Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
           Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
           Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
           Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
           Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
           Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
           Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
           Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
           Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
           Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
           Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
           Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
           Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
           Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
           Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
           Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
           Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
           Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
           Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
           Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
           Rogers,J., Birney,E. and Hayashizaki,Y.
CONSRTM   FANTOM Consortium
TITLE   Analysis of the mouse transcriptome based on functional annotation
           of 60,770 full-length cDNAs
JOURNAL   Nature 420 (6915), 563-573 (2002)
PUBMED   12466851
REFERENCE 6
AUTHORS   Carninci,P., Kasukawa,T., Katayama,S., Gough,J., Frith,M.C.,
           Maeda,N., Oyama,R., Ravasi,T., Lenhard,B., Wells,C., Kodzius,R.,
           Shimokawa,K., Bajic,V.B., Brenner,S.E., Batalov,S., Forrest,A.R.,
           Zavolan,M., Davis,M.J., Wilming,L.G., Aidinis,V., Allen,J.E.,
           Ambesi-Impiombato,A., Apweiler,R., Aturaliya,R.N., Bailey,T.L.,
           Bansal,M., Baxter,L., Beisel,K.W., Bersano,T., Bono,H., Chalk,A.M.,
           Chiu,K.P., Choudhary,V., Christoffels,A., Clutterbuck,D.R.,
           Crowe,M.L., Dalla,E., Dalrymple,B.P., de Bono,B., Della Gatta,G.,
           di Bernardo,D., Down,T., Engstrom,P., Fagiolini,M., Faulkner,G.,
           Fletcher,C.F., Fukushima,T., Furuno,M., Futaki,S., Gariboldi,M.,
           Georgii-Hemming,P., Gingeras,T.R., Gojobori,T., Green,R.E.,
           Gustincich,S., Harbers,M., Hayashi,Y., Hensch,T.K., Hirokawa,N.,
           Hill,D., Huminiecki,L., Iacono,M., Ikeo,K., Iwama,A., Ishikawa,T.,
           Jakt,M., Kanapin,A., Katoh,M., Kawasawa,Y., Kelso,J., Kitamura,H.,
           Kitano,H., Kollias,G., Krishnan,S.P., Kruger,A., Kummerfeld,S.K.,
           Kurochkin,I.V., Lareau,L.F., Lazarevic,D., Lipovich,L., Liu,J.,
           Liuni,S., McWilliam,S., Madan Babu,M., Madera,M., Marchionni,L.,
           Matsuda,H., Matsuzawa,S., Miki,H., Mignone,F., Miyake,S.,

Morris,K., Mottagui-Taba...S., Mulder,N., Nakano,N., Nakauchi,H.,
Ng,P., Nilsson,R., Nishi...uchi,S., Nishikawa,S., Nori,F., Ohara,O.,
Okazaki,Y., Orlando,V., ...ang,K.C., Pavan,W.J., Pavesi,G.,
Pesole,G., Petrovsky,N., ...Piazza,S., Reed,J., Reid,B.Z., Ring,B.Z.,
Ringwald,M., Rost,B., Ru...n,Y., Salzberg,S.L., Sandelin,A.,
Schneider,C., Schonbach, ..., Sekiguchi,K., Semple,C.A., Seno,S.,
Sessa,L., Sheng,Y., Shib...ta,Y., Shimada,H., Shimada,K., Silva,D.,
Sinclair,B., Sperling,S. ...Stupka,E., Sugiura,K., Sultana,R.,
Takenaka,Y., Taki,K., Ta...moja,K., Tan,S.L., Tang,S., Taylor,M.S.,
Tegner,J., Teichmann,S.A ...Ueda,H.R., van Nimwegen,E., Verardo,R.,
Wei,C.L., Yagi,K., Yaman ...shi,H., Zabarovsky,E., Zhu,S., Zimmer,A.,
Hide,W., Bult,C., Grimmo ...d,S.M., Teasdale,R.D., Liu,E.T.,
Brusic,V., Quackenbush,J ...Wahlestedt,C., Mattick,J.S., Hume,D.A.,
Kai,C., Sasaki,D., Tomar ...Y., Fukuda,S., Kanamori-Katayama,M.,
Suzuki,M., Aoki,J., Arak ...awa,T., Iida,J., Imamura,K., Itoh,M.,
Kato,T., Kawaji,H., Kawa ...ashira,N., Kawashima,T., Kojima,M.,
Kondo,S., Konno,H., Naka ...o,K., Ninomiya,N., Nishio,T., Okada,M.,
Plessy,C., Shibata,K., S...hiraki,T., Suzuki,S., Tagami,M., Waki,K.,
Watahiki,A., Okamura-Oho ...Y., Suzuki,H., Kawai,J. and Hayashizaki,Y.
CONSRTM   FANTOM Consortium
TITLE   The transcriptional lan...cape of the mammalian genome
JOURNAL   Science 309 (5740), 1559...1563 (2005)
PUBMED   16141072
REFERENCE 7
AUTHORS   Katayama,S., Tomaru,Y., ...asukawa,T., Waki,K., Nakanishi,M.,
           Nakamura,M., Nishida,H., ...wap,C.C., Suzuki,M., Kawai,J., Suzuki,H.,
           Carninci,P., Hayashizaki ...Y., Wells,C., Frith,M., Ravasi,T.,
           Pang,K.C., Hallinan,J., ...attick,J., Hume,D.A., Lipovich,L.,
           Batalov,S., Engstrom,P.G ...Mizuno,Y., Faghihi,M.A., Sandelin,A.,
           Chalk,A.M., Mottagui-Tab ...r,S., Liang,Z., Lenhard,B. and
           Wahlestedt,C.
CONSRTM   RIKEN Genome Exploration ...Research Group
TITLE   Antisense transcription ...n the mammalian transcriptome
JOURNAL   Science 309 (5740), 1564...1566 (2005)
PUBMED   16141073
REFERENCE 8 (bases 1 to 3222)
AUTHORS   Arakawa,T., Carninci,P., ...Fukuda,S., Hashizume,W., Hayashida,K.,
           Hori,F., Iida,J., Imamur ...K., Imotani,K., Itoh,M., Kanagawa,S.,
           Kawai,J., Kojima,M., Kon...o,H., Murata,M., Nakamura,M., Ninomiya,N.,
           Nishiyori,H., Nomura,N., ...Ohno,M., Sakazume,M., Sano,H., Sasaki,D.,
           Shibata,K., Shiraki,T., ...agami,M., Tagami,Y., Waki,K., Watahiki,A.,
           Muramatsu,M. and Hayashi ...aki,Y.
TITLE   Direct Submission
JOURNAL   Submitted (14-APR-2004) ...oshihide Hayashizaki, The Institute of
           Physical and Chemical Re...earch (RIKEN), Laboratory for Genome
           Exploration Research Gro...p, RIKEN Genomic Sciences Center (GSC),
           RIKEN Yokohama Institute ...1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
           Kanagawa, 230-0045, Japa...(E-mail:genome-res@gsc.riken.jp,
           URL:http://genome.gsc.ri...en.jp/, Tel:81-45-503-9222,
           Fax:81-45-503-9216)
COMMENT   cDNA library was prepare ...and sequenced in Mouse Genome
           Encyclopedia Project of ...enome Exploration Research Group in RIKEN.
           Genomic Sciences Center ...nd Genome Science Laboratory in RIKEN.
           Division of Experimental ...Animal Research in Riken contributed to
           prepare mouse tissues.
           Tissues were provided by ...Dr. John Todd (Dept. of Medical Genetics
           Wellcome Trust Centre fo...Molecular Mechanisms in Disease Wellcome
           Trust/MRC building Adden...rookes Hospital Cambridge) whose
           assistance we gratefully ...acknowledge.
           Please visit our web sit...for further details.
           URL:http://genome.gsc.ri...en.jp/
           URL:http://fantom.gsc.ri...en.jp/.
FEATURES   Location/Qualif...ers
     source   1..3222
                     /organism="Mus ...usculus"
                     /mol_type="mRNA...
                     /strain="NOD"
                     /db_xref="FANTO...DB:F830206C17"
                     /db_xref="taxon...10090"
                     /clone="F830206...C17"
                     /tissue_type="a...tivated spleen"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
     CDS   142..1359

```
/note="unnamed protein product; putative
sialyltransferase 8 (alpha-2, 8-sialyltransferase) F
(MGD|MGI:2386797 GB|NM_145838, evidence: BLASTN, 98%,
match=3156)"
/codon_start=1
/protein_id="BAF42986.1"
/db_xref="GI:74186461"
/translation="MRSGGTLFALIGSLMLLLLLRMLWCPADAPARSRLLMEGSREDT
SGTSAALKTLMSPTTPVPRTRNSTYLDEKTTQITEKCKDLQYSLNSLSNKTRRYSEDD
YLQTITNIQRCPWNRQAEBYDNFRAKLASCCDAIQDFVVSQNNTPVGTNMSYEVESKK
HIPIRENIPHMPVSQPPVDYPYNQCAVVGNGGILNKSLCGAEIDKSDFVFRCNLPPI
TGSASKDVGSKTNLVTVNPSIITLKYQNLKEKKAQPLEDISTYGDAFLLLPAFSYRAN
TGISFKVYQTLKESKMRQKVLFFHPRYLRHLALFWRIKGVTAYRLSTGLMIASVAVEL
CENVKLYGFWPFSKTIEDTPLSHHYYDNMLPKHGFHQMPKEYSQMLQLHRGILLKLQF"

Query Match       96.9%;  Score 3068.6;  DB 6;  Length 3222;
Best Local Similarity  98.7%;  Pred. No. 0;
Matches 3115;  Conservative 0;  Mismatches 39;  Indels 3;  Gaps 2;

Qy    1  CGGAGCGGCGAGTCGGTGCCGCCCGGCGGCTGCGCCCCGGCAGCTTGGCGGCGGAGG    60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   66  CGGAGCGGCGAGTCGGTGCCGCCCGGCGGCTGCGCCCCGGCAGCTTGGCGGCGGAGG   125

Qy   61  ACGCCCGTGGCTCAGGATGAGATCGGGGGGCACGCTGTTCGCCCTCATAGGCAGCCTGAT  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  126  ACGCCCGTGGCTCAGGATGAGATCGGGGGGCACGCTGTTCGCCCTCATAGGCAGCCTGAT  185

Qy  121  GCTGCTGCTCCTCCTGCGTATGCTCTGGTGCCCAGCCGACGCGCCTGCCGCTCCAGCCT  180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  186  GCTGCTGCTCCTCCTGCGTATGCTCTGGTGCCCAGCCGACGCGCCTGCCGCTCCAGCCT  245

Qy  181  GTTGATGGAGGGAAGCAGAGAGACAGAGGACACCAGTGTACCTGAGCTGCACTGAAGACACTCTG  240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  246  GTTGATGGAGGGAAGCAGAGAGACAGAGGACACCAGTGTACCTGAGCTGCACTGAAGACACTCTG  305

Qy  241  GAGCCCGACAACCCCGTACCACGCACCAGGAACAGCACATATCTGGATGAGAAGAACAAC  300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  306  GAGCCCGACAACCCCGTACCACGCACCAGGAACAGCACATATCTGGATGAGAAGAACAAC  365

Qy  301  CCAAATAACAGAGAAGATCTGCAAATATGCTTGAACTCTTATCTTAACAAAAC  360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  366  CCAAATAACAGAGAAGAATGCAAATACCTGCAATATAGCTTGAACTCTTTATCTTAACAAAAC  425

Qy  361  GAGACGGTACTCGAGGATGACTACCTCCAGAACACAGAGATGCCCATG  420
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  426  GAGACGGTACTCGAGGATGACTACCTCCAGAACACAGAGATGCCCATG  485

Qy  421  GAACCGGCAAGCAGAAGAATATGACAATTTAGAGACAAAACTGGCTTCCTGTTCCTATGTGC  480
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  486  GAACCGGCAAGCAGAAGAATATGACAATTTAGAGACAAAACTGGCTTCCTGTTCCTATGGC  545

Qy  481  CATTCAAGACTTCGTGGTTTCCAGAACAGCTACGA  540
         |||||||||||||||||||||||||||||||||||
Db  546  CATTCAAGACTTCGTGGTTTCCAGAACACTCCAGTGGGGACTAACATGAGCTACGA  605

Qy  541  GGTGGAAGCAGAAGAAACACATCCCATTCGAGAGAACATTTTCCACATGTTTCCAGTGTC  600
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  606  GGTGGAAGCAGAAGAAACACATCCCATTCGAGAGAACATTTTCCACATGTTTCCAGTGTC  665

Qy  601  GCAGCCTTTGTGGACTATCCCTATAACCAGTGTGCAGTGGTTGGTAATGGGGAATTCT  660
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  666  ACAGCCCTTTGTGGACTATCCCTATAACCAGTGTGCAGTGGTTGGTAATGGGGAATTCT  725

Qy  661  CAACAAGTCTCTGCGGAGCAGAAATTGATAAATCTGACTTCGTCTTCAGGTGTAACCT  720
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  726  CAACAAGTCTCTGCGGAGCAGAAATTGATAAATCTGACTTCGTCTTCAGGTGTAACCT  785

Qy  721  CCCCCAATCACAGGGAGCGCTAGTAAAGATGTTGGAAGCAAAACAAATCTTGTGACTGT  780
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  786  CCCCCAATCACAGGGAGCGCTAGTAAAGATGTTGGAAGCAAAACAAATCTTGTGACTGT  845

Qy  781  CAATCCCAGCATTATAACCCTGAAGTACCAGAATTTGAAGGAGCAGAAGCACAGTTTTT  840
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  905  CAATCCCAGCATTATAACCCTG...

Db  841  GGAGGACAATCTCCACCTATGGAATGCATTCCTCCTGCTGCCAGCATTTTCCTATCGGGC   900

Qy  901  CAACACAGGCATCTCTTTAAAATCTACCAAACACTCAAAGAGTCAAAAATGAGGCAAAA   960
         ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db  966  CAACCAAACACTCAAAGAGTCAAAAATGAGGCAAAA

Qy  961  GGTTCTCTTCTTCATCCCAGGACCTGAGACACCTCGCTCTTTTCTGGAGAACTAAAGG  1020
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1026  GGTTCTCTTCTTCACCCCAGGACCTGAGACACCTCGCTCTTTTCTGGAGAACTAAAGG  1085

Qy 1021  CAGGCTTGATGATTGCAAGTGTCGCTGTGGGAACTGTG  1080
         ||||||||||||||||||||||||||||||||||||||
Db 1086  GGTGACTGCATACCGCTTGTCC...CAGGCTTGATGATTGCAAGTGTCGCTGTGGGAACTGTG  1145

Qy 1081  TGAAAACGTGAAGCTCACGGATCTCGGCCTTTCCTAAGACTATCGAAGACACCCACT  1140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1146  TGAAAACGTGAAGCTCACGGATCTCGGCCTTTCTCTAAGACTATCGAAGACACCCACT  1205

Qy 1141  CAGTCACCACTACTATGATAACTGTTACCTAAGCATGGTTTCCACCAGATGCCTAAAGA  1200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1206  CAGTCACCACTACTATGATAACTGTTACCTAAGCATGGTTTCCACCAGATGCCTAAAGA  1265

Qy 1201  ATACAGCCAAATGCTCCAGCTCATATGAGAGGAATCCTCAAACTGCAATTCAGCAAATG  1260
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1266  ATACAGCCAAATGCTCCAGCTCATATGAGAGGAATCCTCAAACTGCAATTCAGCAAATG  1325

Qy 1261  TGAAACGGCTTAACGTTCTTAAAGGAGGAGAATAATTCAGGAGGTGGAGTTGGATGGTGTCA  1320
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1326  TGAAACGGCTTACCGTTTCTTAAAGGAGGAATAATTCAGGAGGTGGAGTTGGATGGTGTCA  1385

Qy 1321  CAGCACATCTCCAAAAAGCCAATAAAGAGGCACAGAGAAAGCATGAAGTTACAAAGGCGCT  1380
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1386  CAGCACATCTCCAAAAAGCCAATAAAGAGGCACAGAGAAAGCATGAAGTTACAAAGGCGCT  1445

Qy 1381  CTCCCACTTGTCTGAGACCAAAGCACCCGCCCCCCACTTTGCAGCCTCCACTTTGCAGCCTCCACGAGTCA  1440
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1446  CTCCCACTTGTCTGAGACCAAAGCACCCGCCCCCCACTTTGCAGCCTCCACGAGTCA  1505

Qy 1441  CTCATTCTCACCTTCAAGTCTTTCTCTGAGAATAGAGACCAAAACATCAGACTTGGAT  1500
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1506  CTCATTCTCACCTTCAAGTCTTTCTCTGAGAATGGAGACCGAAACATCAGACTTGGAT  1565

Qy 1501  AAGTAAAATGAGATAAATTTTCAATCATCATGAAATTTGATTTGAGCC--AGGGTCTCT  1558
         ||||||||||||||||||||||||||||||||||||||||||||||||  |||||||||
Db 1566  AATCATCATGAAATTTTCAATCATCATGAAATTTGATTTGAGATTTGAGTTTGAGTTTGAGTTTGAGTTTGAGATTTGATTTGATTTGAT  1625

Qy 1559  CAGAATGTCTTCCTTGTTCCTATCATGATAGCCATTCCCACCTTTATCAGAGTGTAATG  1618
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1626  CAGAATGTCTTCCTTGTTCCTATCATGATAGCCATTCCCACCTTTATCAGAGTGTAATG  1685

Qy 1619  AAACTGTGCAATTGTGCCAAAGCCCTTTCTGAAGAGAATGTCTGAATCATGCGCCGAGT  1678
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1686  AAACTGTGCAATTGTGCCAAAGCCCTTTCTGAAGAGAATGTCTGAATCATGCCCCGAGT  1745

Qy 1679  TTTTACACACAGCTCTTCCTTTAAATAAATCCTTCCCATTCTCCCTCCTCCCTAGAGTA  1738
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1746  TTTTACACACAGCTCTTCCTTTAAATAAATCCTTCCCATTCTCCCTCCTCCCTAGAGTA  1805

Qy 1739  CAGAAACAAAATACCCTGATGTTCAGGAAGAAAAAGTCTTTTTTACTTAGCAATGTGCC  1798
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1806  CAGAAACAAAATACCCTGATGTTCAGGAAGAAAAAGTCTTTTTACTTAGCAATGTGCC  1865

Qy 1799  TGCTTCTGATTCAGTTCGTCTTGGACATTAAGCTGGGTTTGGTTGGATTTGGG  1858
         |||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db 1866  TGCTTCTGATTCAGTTCGTCTTGGACGT-AGCTGGGTTTGGGGTTTGGATTTGTA  1924

Qy 1859  GCGTTTCTTCACTTCTTTTGTCATATTTTCCTTACTTTATCAGTTTGTATTCGAGCTT  1918
         |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db 1925  GCGTTTCTTCACTTCTTTTGTCATATTTTCCTTACTTTATCAGTTTGTACTCGAGCTT  1984
```

Qy 2999 ATATTTTGCAAAGGGTAAAAGG...GTTTTGTTAAATAAATAAAATTATTATTTCTTC 3058
Db 3065 ATATTTTGCAAAGGGTAAAAGG...TTTTGTTAAATAAATAAAATTATCATTTTCTTC 3124
Qy 3059 TGATGAATAGAGGCTCTTTAT...GTGCTGCTAATGAACCTAATTAGCTTTAAATTATCTC 3118
Db 3125 TGATGAATAGAGGCTCTTTAT...GTGCTGCTAATGAACCTAATTAGCTTTAAATAATCTC 3184
Qy 3119 CTAGCAACATTGGTCACGTTC...CATCATGCTATTAGC 3155
Db 3185 CTAGCAACATTGGTCACGTTC...CATCATGCTATTAGC 3221

RESULT 2
AK156557
LOCUS   AK156557   695 bp   mRNA   linear   HTC 21-SEP-2005
DEFINITION   Mus musculus activated spleen cDNA, RIKEN full-length enriched
             library, clone:F830029M2 product:sialyltransferase 8 (alpha-2,
             8-sialyltransferase) F, full insert sequence.
ACCESSION   AK156557
VERSION   AK156557.1 GI:74224371
KEYWORDS   HTC; CAP trapper.
SOURCE   Mus musculus (house mouse)
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
             Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE   High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-4 (1999)
  PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
             Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to
             prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
             Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
             Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
             Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
             Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
             Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
             Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE   RIKEN integrated sequence analysis (RISA) system-384-format
             sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
             Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
             Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
             Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
             Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
             Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
             Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
             Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
             Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
             Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
             Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
             Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
             Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
             Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
             Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
             Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
             Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
             Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
             and Hayashizaki,Y.
CONSRTM   RIKEN Genome Exploration Research Group Phase II Team and the
             FANTOM Consortium

Qy 1919 CCTGCTTGGGATTCTGCAATTCTCTCCCACTGACAGGATCAACTCAATGACATAAAG 1978
Db 1985 CCTGCTTGGGATTCTGCAATTCTCTCCCACTGACAGGATCAACTCAATGACATAAAG 2044
Qy 1979 TAGTTCAAACATCCATTGCTTCTCCACATGTTTATCCATAAAGTTACTCATCTGATTTTA 2038
Db 2045 TAGCTCAAACATCCACTGCTTCTCCACATGTTTATCCATAAAGTTACTCATCTGATTTTA 2104
Qy 2039 TTTAAAATAGTGAACATCTACTTGATATCAGACCCGAGGACCATCCTCCATTGGAGATA 2098
Db 2105 TTTAAAATAGTGAACATCTACTTGATATCAGACCCGAGGACCGTCCTCCATTGGAGATA 2164
Qy 2099 TGAAGATATTGTCACTGCAGAAAGCAGGTGTGTGCCATTAATTGATAAGATACCACAA 2158
Db 2165 TGAAGATATTGTCACTGCAGAAAGCAGGTGTGTGCCATTAATTGATAAGATACCACAA 2224
Qy 2159 GCATCATCATGCCAGTTATGAACACAGTGCTGAAAGGATCATAGACAGGGGTGGTTAAAT 2218
Db 2225 GCATCATCATGCCAGTTATGAACACAGTGCTGAAAGGATCATAGACAGGGGTGGTTAAAT 2284
Qy 2219 CTGATCCCAGTAGAATAAACTTCAGTGTACCTATTTCAGGGAAGAGTTAATTTCACAATT 2278
Db 2285 CTGATCCCAGTAGAATCAACTTCAATGTACCTATTTCAGGGAAGAGTTAATTTCACAATT 2344
Qy 2279 AAAACTAGTAAATGAACCAATTCTTAGGCACATTAAGTGGATTCTGAGTAAAAGAAAGGG 2338
Db 2345 AAAACTAGTAAATGAACCAATTCTTAGGCACATTAAGTGGATTCTGAGTAAAAGAAAGGG 2404
Qy 2339 AACAGCAGGAGAAAGCTGTTCGCTTGGTTCTGATTACCCAAATGAGCATGCTGGAAGGAG 2398
Db 2405 AACAGCAGGAGAAAGCTGTTCGCTTGGTTCTGATTACCCAAATGAGCATGCTGGAAGGAG 2464
Qy 2399 GTTGTGAGGCTACGCTAAAACCTCGCGTAGGAGAGTACAGTGCATGAGTGTGGGGG 2458
Db 2465 GTTGTGAGGCTACGCTAAAACCTCGTGTGTAGGAGAGTACAGTGCATGAGTGTGACGG 2524
Qy 2459 CTTTTGTTGTTCCACACTCGTGAAGGTGAGTAATTCAGAGCCAATCACATCACAAGGATGGAC 2518
Db 2525 CTTTTGTTGTTCCACACTCGTGAAGGTGAGTAATTCAGAGCCAATCACATCACAAGGATGGAC 2584
Qy 2519 ACACCTAACTACTCATCACTTCAGGGGGAGATGAATGCTTTCATGAGAAATTACACTCATAAG 2578
Db 2585 ACACCTAACTACTCATCACTTCAGGGGGAGATGAACGCTTTCATGAGAAATTACACTCATAAG 2644
Qy 2579 CTAAGCATCAGTTTTGAGTAAATTTGAGTAGATGTTAAATATGAACATTTTATACCTCT 2638
Db 2645 CTAAGCATCAGTTTTGAGTAAATTTGAGTAGATGTTAAATATGATCATTTTACACCTCT 2704
Qy 2639 TACTAATGTCCCACCGACACCTTTAATGTAAGCACATTTATTATTAAGTTACTTGACA 2698
Db 2705 TACTAATGTCTCACTGACCACCCTTTTAATGTAAGCACACATTTATTAAGTTAAGTTATTTGACA 2764
Qy 2699 TTAAATGCTTATGTCTGTATATTCTGTTCATCCATCCGATTTTCCCAAAAGTAAGAGCAT 2758
Db 2765 TTACAAGCTTATGTCTGTATATTCGTTCATCCATCCATCAATTTTCCAAAAGTAAGAGCAT 2824
Qy 2759 AGGAGATGAGGCCTACACTGCCAAGAAAACTATAAATTTACTCTTTAATTCTTACTTGAG 2818
Db 2825 AGGAGATGAGGCCCACATGCCAAGAAAACTATAAATTTACTGTTTAATTCTTACTTGAG 2884
Qy 2819 CCAGCTTGTTGTTTATCAAGTGCTTTTTGAAGAGACAGCACCCTGTGAATTCTTCATTC 2878
Db 2885 CCAGCTTGTTGTTTATCAAGTGCTTTTTGAAGAGACAGCACCCCTGTTTAATTCTTCATTC 2944
Qy 2879 TGATACAGTGTCACCTTGTATTTAACATTTGTAATGTTGTTTCAAGTTTACATCTCTTTC 2938
Db 2945 TGATACAATGTCACCTTGTATTTAACATTTGTAATGTTGTTTCAAGTTTATATCTCTTTC 3004
Qy 2939 ATTCTTTTTATAGCAAATCAAACTGATTAGCTTCAGAAATTTATCAGAAGTTCATATATAA 2998
Db 3005 ATTCTTTTTATAGCAAATCAAACTGATTAGCTTCAGAAATTTATCAGAAGTTCATATATAA 3064

TITLE   Functional annotation of a full-length mouse cDNA collection
JOURNAL   Nature 409 (6821), 685-690 (2001)
PUBMED   11217851
REFERENCE   5
AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
            Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
            Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
            Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
            Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
            Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
            Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
            Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
            Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
            Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
            Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
            Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
            Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
            Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
            Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
            Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
            Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
            Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
            Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
            Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
            Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
            Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
            Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
            Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
            Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
            Rogers,J., Birney,E. and Hayashizaki,Y.
CONSRTM   FANTOM Consortium
TITLE   Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
JOURNAL   Nature 420 (6915), 563-573 (2002)
PUBMED   12466851
REFERENCE   6
AUTHORS   Carninci,P., Kasukawa,T., Katayama,S., Gough,J., Frith,M.C.,
            Maeda,N., Oyama,R., Ravasi,T., Lenhard,B., Wells,C., Kodzius,R.,
            Shimokawa,K., Bajic,V.B., Brenner,S.E., Batalov,S., Forrest,A.R.,
            Zavolan,M., Davis,M.J., Wilming,L.G., Aidinis,V., Allen,J.E.,
            Ambesi-Impiombato,A., Apweiler,R., Aturaliya,R.N., Bailey,T.L.,
            Bansal,M., Baxter,L., Beisel,K.W., Bersano,T., Bono,H., Chalk,A.M.,
            Chiu,K.P., Choudhary,V., Christoffels,A., Clutterbuck,D.R.,
            Crowe,M.L., Dalla,E., Dalrymple,B.P., de Bono,B., Della Gatta,G.,
            di Bernardo,D., Down,T., Engstrom,P., Fagiolini,M., Faulkner,G.,
            Fletcher,C.F., Fukushima,T., Furuno,M., Futaki,S., Gariboldi,M.,
            Georgii-Hemming,P., Gingeras,T.R., Gojobori,T., Green,R.E.,
            Gustincich,S., Harbers,M., Hayashi,Y., Hensch,T.K., Hirokawa,N.,
            Hill,D., Huminiecki,L., Iacono,M., Ikeo,K., Iwama,A., Ishikawa,T.,
            Jakt,M., Kanapin,A., Katoh,M., Kawasawa,Y., Kelso,J., Kitamura,H.,
            Kitano,H., Kollias,G., Krishnan,S.P., Kruger,A., Kummerfeld,S.K.,
            Kurochkin,I.V., Lareau,L.F., Lazarevic,D., Lipovich,L., Liu,J.,
            Liuni,S., McWilliam,S., Madan Babu,M., Madera,M., Marchionni,L.,
            Matsuda,H., Matsuzawa,S., Miki,H., Mignone,F., Miyake,S.,
            Morris,K., Mottagui-Tabar,S., Mulder,N., Nakano,N., Nakauchi,H.,
            Ng,P., Nilsson,D., Nishiguchi,S., Nishikawa,S., Nori,F., Ohara,O.,
            Okazaki,Y., Orlando,V., Pang,K.C., Pavan,W.J., Pavesi,G.,
            Pesole,G., Petrovsky,N., Piazza,S., Reed,J., Reid,J.F., Ring,B.Z.,
            Ringwald,M., Rost,B., Ruan,Y., Salzberg,S.L., Sandelin,A.,
            Schneider,C., Schonbach,C., Sekiguchi,K., Semple,C.A., Seno,S.,
            Sessa,L., Sheng,Y., Shibata,Y., Shimada,H., Shimada,K., Silva,D.,
            Sinclair,B., Sperling,S., Stupka,E., Sugiura,K., Sultana,R.,
            Takenaka,Y., Taki,K., Tammoja,K., Tan,S.L., Tang,S., Taylor,M.S.,
            Tegner,J., Teichmann,S.A., Ueda,H.R., van Nimwegen,E., Verardo,R.,
            Wei,C.L., Yagi,K., Yamanishi,H., Zabarovsky,E., Zhu,S., Zimmer,A.,
            Hide,W., Bult,C., Grimmond,S.M., Teasdale,R.D., Liu,E.T.,
            Brusic,V., Quackenbush,J., Wahlestedt,C., Mattick,J.S., Hume,D.A.,
            Kai,C., Sasaki,D., Tomaru,Y., Fukuda,S., Kanamori-Katayama,M.,
            Suzuki,M., Aoki,J., Arakawa,T., Iida,J., Imamura,K., Itoh,M.,
            Kato,T., Kawaji,H., Kawagashira,N., Kawashima,T., Kojima,M.,
            Kondo,S., Konno,H., Nakano,K., Ninomiya,N., Nishio,T., Okada,M.,
            Plessy,C., Shibata,K., Shiraki,T., Suzuki,S., Tagami,M., Waki,K.,

CONSRTM   Watahiki,A., Okamura-Oho,Y., Suzuki,H., Kawai,J. and Hayashizaki,Y.
            FANTOM Consortium
TITLE   The transcriptional landscape of the mammalian genome
JOURNAL   Science 309 (5740), 1559-1563 (2005)
PUBMED   16141072
REFERENCE   7
AUTHORS   Katayama,S., Tomaru,Y., Kasukawa,T., Waki,K., Nakanishi,M.,
            Nakamura,M., Nishida,H., Yap,C.C., Suzuki,M., Kawai,J., Suzuki,H.,
            Carninci,P., Hayashizaki,Y., Wells,C., Frith,M., Ravasi,T.,
            Pang,K.C., Hallinan,J., Mattick,J., Hume,D.A., Lipovich,L.,
            Batalov,S., Engstrom,P.G., Mizuno,Y., Faghihi,M.A., Sandelin,A.,
            Chalk,A.M., Mottagui-Tabar,S., Liang,Z., Lenhard,B. and
            Wahlestedt,C.
CONSRTM   RIKEN Genome Exploration Research Group
TITLE   Antisense transcription in the mammalian transcriptome
JOURNAL   Science 309 (5740), 1564-1566 (2005)
PUBMED   16141073
REFERENCE   8 (bases 1 to 3695)
AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Hashizume,W., Hayashida,K.,
            Hori,F., Iida,J., Imamura,K., Imotani,K., Itoh,M., Kanagawa,S.,
            Kawai,J., Kojima,M., Konno,H., Murata,M., Nakamura,M., Ninomiya,N.,
            Nishiyori,H., Nomura,N., Ohno,M., Sakazume,N., Sano,H., Sasaki,D.,
            Shibata,K., Shiraki,T., Tagami,M., Tagami,Y., Waki,K., Watahiki,A.,
            Muramatsu,M. and Hayashizaki,Y.
TITLE   Direct Submission
JOURNAL   Submitted (30-MAR-2004) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa, 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
            URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT   and sequenced in Mouse Genome
            cDNA library was prepared
            Encyclopedia Project of the
            Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
            Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
            Trust/MRC building Addenbrookes Hospital Cambridge) whose
            assistance we gratefully acknowledge.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.jp/.
            URL:http://fantom.gsc.riken.jp/.
FEATURES   Location/Qualifiers
     source   1..3695
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="NOD"
                     /db_xref="taxon:10090"
                     /clone="F830029M21"
                     /tissue_type="activated spleen"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
     misc_feature   1..3695
                     /note="putative
                     sialyltransferase 8 (alpha-2, 8-sialyltransferase) F
                     [MGD|MGI:238679 GB|NM_145838, evidence: BLASTN, 98%,
                     match=3160)"
ORIGIN

Query Match   95.6%; Score 3026.2; DB 6; Length 3695;
Best Local Similarity 98.6%; Prd. No. 0;
Matches 3116; Conservative 0; Mismatches 38; Indels 7; Gaps 6;

Qy   1 CGGAGCGGCGGAGTCGGTGCCGC-CCCCGGCAGCTTTGGCGGCGAG 59
           |||||||||||||||||||||||| ||||||||||||||||||||
Db   536 CGGAGCGGCGGAGTCGGTGCCGCCGGGGCGGCAGCTTTGGCGGCGAG 595

Qy   60 GACGCCCCGGGCTCAGGATGAG-T-CGGGGGGCACGCGTTCGCCCTCATAGGAGCCTG 118
           |||||||||||||||||||||| | |||||||||||||||||||||||||||||||||
Db   596 GACGCCCCGGGCTCAGGATGAGTGTGCGGGGGGCACGCGTTCGCCCTCATATGCAGCCTG 655

```
  119 ATGCTGCTGCTCCTCTGCGTATGCTCTTGGTGTTGCCCAGCGACCGGCCCGCCTCCAGG  178        Qy   1736 AAGATACAGCCAAATGCTCCA CTCCATATGAGGAGGAATCCTCAAACTGCAATTCAGCA 1795
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||                    ||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
  656 ATGCTGCTGCTCCTCTGCGTATGCTCTTGGTGTTGCCCAGCGACCGGCCCGCCTCCAGG  715        Db

  179 CTGTTGATGGAGGAAGCAGAGAGGAGAGAGAGAGACCAGTTGCTACCTCAGCTGCACTC  238        Qy   1257 AATGTGAAACGGCTAACGTTT TTAGAAGGAGAATAATTCAGGAGGTGGAGTGGAGTGT  1316
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||                    ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
  716 CTGTTGATGGAGGAAGCAGAGAGGAGAGAGAGAGACCAGTTGCTACCTCAGCTGCACTC  775        Db

  239 TGGAGCCGACAACCCCGGTACCACGCGACCAGGAACAGCACATATCTGGATGAGAAGACA 298        Qy   1796 AATGTGAAACGGCTTAACGTTT TTAGAAGGAGAATAATTCAGGAGGTGGAGTGGAGTGT 1855
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||                    ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
  776 TGGAGCCGACAACCCCGGTACCACGCGACCAGGAACAGCACATATCTGGATGAGAAGACA 835        Db

  299 ACCCAAATAACAGAGAAGATCGCAAAGATCTGCAATATAGCTTGAACTCTTTATCTAACAAA 358    Qy   1317 GTCACAGCATCTCCAAAAGCC CATAGAAGAAGGCACAGAGAAAGCATGAATTACAAAGG 1376
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||                    ||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
  836 ACCCAAATAACAGAGAAAGATGCAAAGATCTGCAATATAGCTTGAACTCTTTATCTAACAAA 895    Db

  359 ACGAGACGGTACTCTGAGGATGACTACCTCCAGACCATCACAAACATACAGAGATGCCCA 418    Qy   1377 CGCTCTCCCACTTGTCTAGACC AAGCCACCCGCCCCACTCACTTGCGAGCCTCCACGA 1436
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||                    ||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
  896 ACGAGACGGTACTCTGAGGATGACTACCTCCAGACCATCACAAACATACAGAGATGCCCA 955    Db

  419 TGGAACCGGCAAGCAGAGAATGACAATTTTAGAGCAAAACTGGCTTCCTGTTGCGAT  478    Qy   1437 GTCACTCATTCTCACCTTCAAC TTCTTTCTGAGAATAGAGACCAAAACATCAGACTT 1496
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||                    ||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
  956 TGGAACCGGCAAGCAGAGAATGACAATTTTAGAGCAAAACTGGCTTCCTGTTGCGAT  1015   Db
```

(The remainder of the page continues as additional DNA sequence alignment blocks between query (Qy) and database (Db) sequences, with position numbers ranging up to 2334 for the upper/right alignment and 1735/1256 for the lower/left alignment.)

```
Db   2815 AATTAAAACTAGTAAATGAACCAATTCTTAGGCACATTAAGTGGATTCTGAGTAAAGAA 2874

Qy   2335 AGGGAACAGCAGGAGGAGAAAGCTGTTCGCTTGGTTCTGATTACCCAAATGAGCATGCTGGAA 2394
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2875 AGGGAACAGCAGGAGGAGAAAGCTGTTCGCTTGGTTCTGATTACCCAAATGAGCATGCTGGAA 2934

Qy   2395 GGAGGTTGTGAGGCTACGCTAAAACCTCTGCGTAGGGAGAGTACAGTGCATGAGTGTG 2454
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2935 GGAGGTTGTGAGGCTACGCTAAAACCTCTGTGTAGGGAGAGTACAGTGCATGAGTGTG 2994

Qy   2455 GCGGCTTTGTTGTCCACACTCGTGAAGGGTGAGTAATTCAGAGCCAATCACATCACAAGGAT 2514
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2995 ACGGGCTTTTGTTGTCCACACTGGTGAAGGGTGAGTAATTCAGAGCCAATCACATCACAAGGAT 3054

Qy   2515 GGACACACCTAACTCACTTCAGGGGGAGGGGAGATGAATGCTTCATGAGAGAATTACACTCA 2574
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3055 GGACACACCTAACTCACTTCAGGGGGAGGGGAGATGAACGCTTTCATGAGAAATTACACTCA 3114

Qy   2575 TAAGTAAGCATCAGTTTTTGAGTAAAATTTGAGTAGATGTTAAATATGAACATTTTTATAC 2634
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3115 TAAGTAAGCATCAGTTTTTGAGTAAAATTTGAGTAGATGTTAAATATGATCATTTTCACAC 3174

Qy   2635 CTCTTACTAATGTCCCACCGACACACCTTTAATGTAAGCACATTTATTTATTAAGTTACTT 2694
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3175 CTCTTACTAATGTCTCACTGACACACCTTTAATGTAAGCACATTTATTTGTTAAGTTATTT 3234

Qy   2695 GACATTAAATGCTTATGTCTGTATATTCTGTTCATCCATCGATTTTCCCAAAAGTAAGA 2754
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3235 GATATTACAAGCTTATGTCTGTATATTCTGTTCATCCATCAATTTTCCCAAAAGTAAGA 3294

Qy   2755 GCATAGGAGATGAGGCCTACATGCCAAGAAACTATAAATTTTACTCTTTAATTCTTACT 2814
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3295 GCATAGGAGATGAGGCCCACATGCCAAGAAACTATAAATTTTACTGTTTAATTCTTACT 3354

Qy   2815 TGAGCCAGCTGTTGTTTATCAAGTGCTTTTTGAAGAGACAGCACCCTGTGAATTCTTC 2874
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3355 TGAGCCAGCTGTTGTTTATCAAGTGCTTTTTGAAGAGACAGCACCCTGTTAATTCTTC 3414

Qy   2875 ATTCTGATACAGTGTCACCTTGTATTTAACATTTGTATTGTTGTTCAAGTTTACATCTC 2934
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3415 ATTCTGATACAATGTCACCTTGTATTTAACATTTGTAATGTTGTTCAAGTTTACATCTC 3474

Qy   2935 TTTCATTCTTTTATAGCAAATCAAACGTATTAGCTTCAGAAATTTATCAGAAGTTCATAT 2994
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3475 TTTCATTCTTTTATAGCAAATCAAACGTATTAGCTTCAGAAATTTATCAGAAGTTCATAT 3534

Qy   2995 ATAAATATTTTGCAAAGGGTAAAAACGCTTTTTTGTTAAATAAAATAAAAATTATTATTTT 3054
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3535 ATAAATATTTTGCAAAGGGTAAAAGGCTTTTTTGTTAAATAAAATAAAAATTATCATTTT 3594

Qy   3055 CCTCTGATGAATAGAGAGGCTCTTTTATGCTGCTAATGAACCTAATTAGCTTTAAATTA 3114
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3595 CCTCTGATGAATATAGGCTCTTTTATGCTGCTAATGAACCTAATTAGCTTTAAATAA 3654

Qy   3115 TCTCCCTAGCCAACATTGGTCACGTTTCAATCATGCTATTAGC 3155
          |||||||||||||||||||||||||||||||||||||||||||
Db   3655 TCTCCCTAGCCAACATTGGTCACGTTTCAATCATGCTATTAGC 3695
```

REFERENCE   1
AUTHORS   Carninci,P. and Hayashizaki,Y.
TITLE   High-efficiency full-length cDNA cloning
JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
PUBMED   10349636
REFERENCE   2
AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
PUBMED   11042159
REFERENCE   3
AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Wataniki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE   RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
PUBMED   11076861
REFERENCE   4
AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
TITLE   Functional annotation of a full-length mouse cDNA collection
JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
AUTHORS   The FANTOM Consortium, the RIKEN Genome Exploration Research Group
            Phase I and II Team.
TITLE   Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6
AUTHORS   RIKEN Genome Exploration Research Group, Genome Science Group
            (Genome Network Core Team) and the FANTOM Consortium.
TITLE   Antisense Transcription in the Mammalian Transcriptome
JOURNAL   Science 309, 1564-1566 (2005)
REFERENCE   7
AUTHORS   The FANTOM Consortium, Riken Genome Exploration Research Group and
            Genome Science Group (Genome Network Project Core Group).
TITLE   The Transcriptional Landscape of the Mammalian Genome
JOURNAL   Science 309, 1559-1563 (2005)
REFERENCE   8 (bases 1 to 3292)
AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Sato,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Toaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
TITLE   Direct Submission
JOURNAL   Submitted (16-APR-2002) Toshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa, 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
            URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT   cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.jp/

```
RESULT 3
AK085105
LOCUS       AK085105                3292 bp    mRNA    linear   HTC 02-SEP-2005
DEFINITION  Mus musculus 13 days embryo lung cDNA, RIKEN full-length enriched
            library, clone:D430037E10 product:weakly similar to
            SIALYLTRANSFERASE 8 (FRAGMENT) [Bos taurus], full insert sequence.
ACCESSION   AK085105
VERSION     AK085105.1  GI:26351460
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
```

```
FEATURES    Location/Qualifiers
source      1..3292
            /organism="Mus musculus"
            /mol_type="mRNA"
            /strain="C57BL/6J"
            /db_xref="FANTOM DB:D430037E10"
            /clone="D430037E10"
            /tissue_type="lung"
            /clone_lib="RIKEN full-length enriched mouse cDNA library"
            /dev_stage="13 days embryo"
CDS         322..1410
            /note="unnamed protein product; putative
            weakly similar to SIALYLTRANSFERASE 8 (FRAGMENT) [Bos
            taurus] (SPTR|AAL47018, evidence: FASTY, 50.3%ID,
            98.7%length, match=489)"
            /codon_start=1
            /protein_id="BAC39367.1"
            /db_xref="GI:26351461"
            /translation="MEGSREDTSGTSAALKTLWSPTTPVPRTRNSTYLDEKTQITEK
            CKDLQYSLNSLSNKTRYSEDDYLQTITNIQRCPWNRQAEEYDNFRAKLASCCDAIQD
            FVVSQNNTPVGTNMSYEVESKKHIPIRENIFHMFPVSQPFVDYPYNQCAVVGNGGILN
            KSLCGAEIDKSDFVFRCNLPPITGSASKDVGSKTNLVTVNPSIIITLKYQNLKEKKAQF
            LEDISTYGDAPLLPAFSYRANTGISFKVYQTLKESKMRQKVLFFHPRYLRHLALFWR
            TKGVTAYRLSTGLMIASVAVELCENVKLYGFWPFSKTIEDTPLSHHYYDNMLPKHGFH
            QMPKEYSQMLQLHMRGILKLQFSKCETA"
ORIGIN
Query Match      95.1%;   Score 3010;   DB 6;   Length 3292;
Best Local Similarity   99.3%;   Pred. No. 0;
Matches 3022;   Conservative   0;   Mismatches  20;   Indels   0;   Gaps   0;

Qy  114 GCCTGATGCTGCTGCTCCTCGCTATGTCCTGGTGTGCCGCCGACGGCGCCTGCCCCGCT 173
        |||||||| | ||||     |    |          |||| |||||||||||||||||||
Db  251 GGCAGCGGCGCCGCGGTGGCCGGCGGAGGAGGAGGAGGAGCCGGTGCCCAGCCGCCCGCT 310

Qy  174 CCAGGCTGTTGATGGAGGAAGCAGAGAGGAGCAGAGGACCAGTGTACCTCAGCTGCACTG 233
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  311 CCAGGCTGTTGATGGAGGAAGCAGAGAGGAGCAGAGGACCAGTGTACCTCAGCTGCACTG 370

Qy  234 CACTCTGGAGCCCGACAACCCCGTACCACGCACCAGGAACAGCACATATCTCGATGAGA 293
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  371 CACTCTGGAGCCCGACAACCCCGTACCACGCACCAGGAACAGCACATATCTCGATGAGA 430

Qy  294 AGACAACCCAAATAACAGAGAAATGCAAAGATCTGCAATATAGCTTGAACTCTTTATCTA 353
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  431 AGACAACCCAAATAACAGAGAAATGCAAAGATCTGCAATATAGCTTGAACTCTTTATCTA 490

Qy  354 ACAAAACGAGACGGTACTCTGAGGATGACTACCTGTCAGCTCAGCTGCACTACAGAGAT 413
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  491 ACAAAACGAGACGGTACTCTGAGGATGACTACCTGTCAGCTCAGCTGCACTACAGAGAT 550

Qy  414 GCCCATGAACCGGCAAGCAGAAGAATATGACAATTTAGAGACAAAACTGGCTTCCTGTT 473
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  551 GCCCATGAACCGGCAAGCAGAAGAATATGACAATTTAGAGACAAAACTGGCTTCCTGTT 610

Qy  474 GCGATGCCATTCAAGACTTCGTGGTTTCCCAGACAACACTCCAGTGGGGACTAACATGA 533
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  611 GCGATGCCATTCAAGACTTCGTGGTTTCCCAGACAACACTCCAGTGGGGACTAACATGA 670

Qy  534 GCTACGAGGTGGAAAGCAAGAACACATCCCATTCGAGAGAACATTTTCCACATGTTTC 593
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  671 GCTACGAGGTGGAAAGCAAGAACACATCCCATTCGAGAGAACATTTTCCACATGTTTC 730

Qy  594 CAGTGTCGCAGCCTTTTGTGGACTATCCCTATAACCAGTGTGCAGTGGTTGGTAATGGGG 653
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  731 CAGTGTCGCAGCCTTTTGTGGACTATCCCTATAACCAGTGTGCAGTGGTTGGTAATGGGG 790

Qy  654 GAATTCTCAACAAGTCTCTCTGCCGAGCAGAGAATTGATAAATCCAGT 713
        ||||||||||||||||||||||||||||||||||||||||||||||||
Db  791 GAATTCTCAACAAGTCTCTCTGCCGAGCAGAGAATTGATAAATCCAGT 850
```

```
Qy  714 GTAACCTCCCCCAATCACAGGCAGCGCTAGTAAAGATGTTGGAGCAAAACAAATCTTG 773
        |||||||||||| ||||||||| |||||||||||||||||||||||||||||||||||
Db  851 GTAACCTCCCCCCAATCACAGGAGCGCTAGTAAAGATGTTGGAGCAAAACAAATCTTG 910

Qy  774 TGACTGTCAATCCCAGCATTATACCCTGAAGTACCAGATTTGAAGGAGGAAGAGACAC 833
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  911 TGACTGTCAATCCCAGCATTATACCCTGAAGTACCAGATTTGAAGGAGGAAGAGACAC 970

Qy  834 AGTTTTTGGAGGAGCATCTCCACTATGGAGATGCATTCCTCCTCCTGCCCAGCATTTCCT 893
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  971 AGTTTTTGGAGGAGCATCTCCACTATGGAGATGCATTCCTCCTCCTGCCCAGCATTTCCT 1030

Qy  894 ATCGGGCCAACACAGGCATCTCTTTAAAGTCTACCAAAACACTCAAAGAGTCAAAAATGA 953
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1031 ATCGGGCCAACACAGGCATCTCTTTAAAGTCTACCAAAACACTCAAAGAGTCAAAAATGA 1090

Qy  954 GGCAAAAGTTCTCTTCTTCCACCCCAGTGACTGACACCTCGCTCTTTTCTCGGAGAA 1013
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1091 GGCAAAAGTTCTCTTCTTCCACCCCAGTGACTGACACCTCGCTCTTTTCTCGGAGAA 1150

Qy 1014 CTAAAGGGGTGACTGCATACGTTGTCCACAGGCTTGATGATTGCAAGTCTCGCTGTGG 1073
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1151 CTAAAGGGGTGACTGCATACGTTGTCCACAGGCTTGATGATTGCAAGTCTCGCTGTGG 1210

Qy 1074 TACGGATTCTGCGCCTTTCTCTAAGACTATCGAAGACATACGGATTCTGCGCCTTTCTCTAAGACTATCGAAGACA 1133
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1211 AACTGTGTGAAAAGTGAAGCTCGGATTCTGCGCCTTTCTCTAAGACTATCGAAGACA 1270

Qy 1134 CCCCACTCAGTCACCACTACTAGATAACATGTTACCTAAGCATGGTTTCCACCAGATGC 1193
        ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db 1271 CCCCACTCAGTCACCACTACTAGATAACATGTTACCTAAGCATGGTTTCCACCAGATGC 1330

Qy 1194 CAGCTCCATATGAGAGGAATCCTCAAACTGCAATTCA 1253
        |||||||||||||||||||||||||||||||||||||
Db 1331 CAGCTCCATATGAGAGGAATCCTCAAACTGCAATTCA 1390

Qy 1254 TTCTTAGAAGGAGAATAATTTCAGGAGGTGGAGTGGA 1313
        |||||||||||||||||||||||||||||||||||||
Db 1391 GCAAATGTGAAACGGCTTAACGTTCTTAGAAGGAGAATAATTTCAGGAGGTGGAGTGGA 1450

Qy 1314 TGTGTCAGCAGCATCCTCCAAAAACCAATAGAAGAAGGCACAGAGAAAGCATGAATTACAA 1373
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1451 TGTGTCAGCAGCATCCTCCAAAAACCAATAGAAGAAGGCACAGAGAAAGCATGAATTACAA 1510

Qy 1374 AGGCGCCTCTCCCCACTTGTCTAGCCAAAGCCACCCGCCCCGCCCCACTCACTTTGCAGCCTCCA 1433
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1511 AGGCGCCTCTCCCCACTTGTCTAGCCAAAGCCACCCGCCCCGCCCCACTCACTTTGCAGCCTCCA 1570

Qy 1434 CGAGTCACTCATTCTCACCTTGCACGTTCTTCTCTGAGAATAGAGACCAAAACATCAGA 1493
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1571 CGAGTCACTCATTCTCACCTTGCACGTTCTTCTCTGAGAATAGAGACCAAAACATCAGA 1630

Qy 1494 CTTGGATAAGTAAAATGATGATTTTTCAAATCATCATAGAATTGATTTGAGCCAGGG 1553
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1631 CTTGGATAAGTAAAATGATGATTTTTCAAATCATCATAGAATTGATTTGAGCCAGGG 1690

Qy 1554 TCTCTCAGATGCTTCCTTGTCTATCCATGATAGCCATTCCCACCTTTATCAGAGTGG 1613
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1691 TCTCTCAGATGCTTCCTTGTCTATCCATGATAGCCATTCCCACCTTTATCAGAGTGG 1750

Qy 1614 TAATGAAACTGTGCAATTGCAAAGACCCTTTCTGAAGAGAATGTCTGAATCATGCGC 1673
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1751 TAATGAAACTGTGCAATTGCAAAGACCCTTTCTGAAGAGAATGTCTGAATCATGCGC 1810

Qy 1674 CTTTATAAATAAATCCTCCCATTCTCCCTCCTAGTA 1733
        ||||||||||||||||||||||||||||||||||||
Db 1811 CTTTATAAATAAATCCTCCCATTCTCCCTCCTAGTA 1870

Qy 1734 GAGTACAGAAACAAAATACCCGATGATTCAGGAAGAAGAAAGTCTTTTTACTTAGCAAT 1793
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1871 GAGTACAGAAACAAAATACCCGATGATTCAGGAAGAAGAAAGTCTTTTTACTTAGCAAT 1930

Qy 1794 GTGTGCCTGCTTCTGATTCAGTCCTTGTGACATTAAGCTGGGTTGGTTGGAT 1853
```

```
Db    1931  GTGCCTGCTTCTGATTCAGTTCGCTTGTGACATTGGGTTTGGTTGGAT  1990
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db    3011  CATTCTGATACAGTGTCACCTTCTATTTAACATTTGTTTGTTTTCAAGTTACATCT  3070

Qy    1854  TCGGGGCCGTTCTTCACTTCTTTTGTCTATATTTTCCTTACCTTTGTATTCG  1913
            |||||||||||||||||||||||||||||||||||||||||||||||||
Qy    2934  CTTTCATTCTTTTATAGCAAATAAACGTATTAGCTTCAGAAATTTATCAGAAGTTCATA  2993

Db    1991  TTGGGGCCGTTCTTCACTTCTTTTGTCTATATTTTCCTTACCTTTGTATTCG  2050
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db    3071  CTTTCATTCTTTTATAGCAAATAAACGTATTAGCTTCAGAAATTTATCAGAAGTTCATA  3130

Qy    1914  AGCTTCCTGCTTTGGGATTCTGCAATTCTCTCTCCCACTGACACGGATCAACTCAATGACA  1973
            |||||||||||||||||||||||||||||||||||||||||||||||||
Qy    2994  TATAAATATTTTGCAAAGGGTAAAGGCTTTTTGTAAATAAAATAAAATTATTATTT  3053

Db    2051  AGCTTCCTGCTTTGGGATTCTGCAATTCTCTCTCCCACTGACACGGATCAACTCAATGACA  2110
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db    3131  TATAAATATTTTGCAAAGGGTAAAGGCTTTTTGTAAATAAAATAAAATTATTATTT  3190

Qy    1974  TAAAGTAGTTCAAACATCCATTGCTTCTCACATGTTTATCCATAAAGTTACTCATCTGA  2033
            |||||||||||||||||||||||||||||||||||||||||||||||||
Qy    3054  TCTTCTGATGAATAGAGGCTCTTTATGCTGCTGCTAATGAACCTAATTAGCTTTAAATT  3113

Db    2111  TAAAGTAGTTCAAACATCCATTGCTTCTCACATGTTTATCCATAAAGTTACTCATCTGA  2170
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db    3191  TCTTCTGATGAATAGAGGCTCTTTATGCTGCTGCTAATGAACCTAATTAGCTTTAAATT  3250

Qy    2034  TTTTATTTAAAAATAGTGAACATCTGATATCAGACCCGAGGACCATCCTCCATTGGA  2093
            |||||||||||||||||||||||||||||||||||||||||||||||||
Qy    3114  ATCCCTAGCAACATTGGTCACTTTCAATCATGCTATTAGC  3155

Db    2171  TTTTATTTAAAAATAGTGAACATCTGATATCAGACCCGAGGACCATCCTCCATTGGA  2230
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db    3251  ATCCCTAGCAACATTGGTCACTTTCAATCATGCTATTAGC  3292

Qy    2094  GAATATGAAGATATTGTCACTGGCAGAAAAGCAGGNGTGTGCCATTAATTGATAAGATAC  2153

Db    2231  GAATATGAAGATATTGTCACTGGCAGAAAAGCAGGNGTGTGCCATTAATTGATAAGATAC  2290

Qy    2154  CACAAGCATCATCATGCCAGTTATGAACACAGTGCTGAAAGGATCATAGAACAGGGGTGT  2213

Db    2291  CACAAGCATCATCATGCCAGTTATGAACACAGTGCTGAAAGGATCATAGAACAGGGGTGT  2350

Qy    2214  TAAATCTGATCCCAGTAGAATAAACTTCAGTGTACCTATTTCAGGGAAGAGTTAATTTCA  2273

Db    2351  TAAATCTGATCCCAGTAGAATAAACTTCAGTGTACCTATTTCAGGGAAGAGTTAATTTCA  2410

Qy    2274  CAATTAAAACTAGTAAATGAACCAATTCCTAGGCACATTAAGTTGGATTCTGAGTAAAAGA  2333

Db    2411  CAATTAAAACTAGTAAATGAACCAATTCCTAGGCACATTAAGTTGGATTCTGAGTAAAAGA  2470

Qy    2334  AAGGGAACAGCAGGAGAAAGCTGTTCGCTTGGTTCTGATTACCCAAATGAGCATGCTGGA  2393

Db    2471  AAGGGAACAGCAGGAGAAAGCTGTTCGCTTGGTTCTGATTACCCAAATGAGCATGCTGGA-  2530

Qy    2394  AGGAGGTTGTGAGGCTACGCTAAAACCTCGCGTAGGAGAGAGTACAGTGCAGTGAGTGT  2453

Db    2531  AGGAGGTTGTGAGGCTACGCTAAAACCTCGCGTAGGAGAGAGTACAGTGCAGTGAGTGT  2590

Qy    2454  GGGGGCTTTGTCCACACTCGTGAAGGGTGAGTAATTCAGAGCCAATCACATCACACAGGA  2513

Db    2591  GGGGGCTTTGTCCACACTCGTGAAGGGTGAGTAATTCAGAGCCAATCACATCACACAGGA  2650

Qy    2514  TGGACACACCTAACTCATCATCACTTCAGGGGGAGATGAATGCTTTCATGAGAAATTACACTC  2573

Db    2651  TGGACACACCTAACTCATCATCACTTCAGGGGGAGATGAATGCTTTCATGAGAAATTACACTC  2710

Qy    2574  ATAAGCTAAGCATCAGTTTTGAGTAAAATTTGAGTAGATGTTAAATATGAACATTTTTATA  2633

Db    2711  ATAAGCTAAGCATCAGTTTTGAGTAAAATTTGAGTAGATGTTAAATATGAACATTTTTATA  2770

Qy    2634  CCTCTTACTAAGTCCCACCGACACCTTTAATGTAAGCACATTTATTTATTAAGTTACT  2693

Db    2771  CCTCTTACTAAGTCCCACCGACACCTTTAATGTAAGCACATTTATTTATTAAGTTACT  2830

Qy    2694  TGACATTAAATGCTTATGTCTGTATATTCTGTTCATCCATCGATTTTCCCAAAAAGTAAG  2753

Db    2831  TGACATTAAATGCTTATGTCTGTATATTCTGTTCATCCATCGATTTTCCCAAAAAGTAAG  2890

Qy    2754  AGCATAGGAGATGAGGCTACATGCCAAGAAAACTATAAAATTTACTCTTAATTCTTAC  2813

Db    2891  AGCATAGGAGATGAGGCTACATGCCAAGAAAACTATAAAATTTACTCTTAATTCTTAC  2950

Qy    2814  TTGAGCCAGCTGTTGTTTATCAAGTGCTTTTTGAAGAGACAGCACCCTGTGAATTCTT  2873

Db    2951  TTGAGCCAGCTGTTGTTTATCAAGTGCTTTTTGAAGAGACAGCACCCTGTGAATTCTT  3010

Qy    2874  CATTCTGATACAGTGTCACCTTGTATTTAACATTTGTTTGTTTTCAAGTTACATCT  2933
```

RESULT 4
CX565877

```
LOCUS       CX565877                        820 bp    mRNA      linear     EST 12-JAN-2005
DEFINITION  UI-M-IB0-cum-m-15-0-UI.r1 NIH_BMAP_IB0 Mus musculus cDNA clone
            IMAGE:30945902 5', mRNA sequence.
ACCESSION   CX565877
VERSION     CX565877.1  GI:57592906
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 820)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. James Lin University of Iowa
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Distribution information can be found at
            http://genome.uiowa.edu/distribution/mousefl.html
            This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1..820
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3945902"
                     /tissue_type="whole eye"
                     /dev_stage="newborn (1,5,15 days)"
                     /lab_host="DH10 (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_IB0"
                     /note="Organ: Eye; Vector: pYX-Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according to
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel.First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site .Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with NotI and then cloned
                     directionally into pYX-Asc vector. The library tag
                     sequence locate between the Not I site and the polyA tail
                     is AATAATTACG. This library was created for the University
                     Iowa Brain Anatomy Project (BMAP). 'Gene Discovery in the
                     Developing Mouse Nervous System', supported by National
                     Institute of Mental Health (NIMH)."
ORIGIN
```

```
Query Match        25.8%;   Score 818; DB 9; Length 820;
Best Local Similarity 99.8%;   Pred. No. 1.8e-200;
Matches 818;  Conservative 0;  Mismatches 2;  Indels 0;  Gaps 0;

Qy  448 TTTTAGAGCAAAACTGGCTTCCTGTGTGCGATGCCATTCAAGACTTCGTGGTTTCCCAGAA 507
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 TTTTAGAGCAAAACTGGCTTCCTGTGTGCGATGCCATTCAAGACTTCGTGGTTTCCCAGAA 60

Qy  508 CAACACTCCAGTGGGGACTAACATGAGCTACGAGGTGAAAGCAAGAAACACATCCCCAT 567
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CAACACTCCAGTGGGGACTAACATGAGCTACGAGGTGAAAGCAAGAAACACATCCCCAT 120

Qy  568 TCGAGAGAACATTTCCACATGTTCCAGTGTCGCAGCCTTTGTGGACTATCCCTATAA 627
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 TCGAGAGAACATTTCCACATGTTCCAGTGTCGCAGCCTTTGTGGACTATCCCTATAA 180

Qy  628 CCAGTGTGCAGTGGTTGGTAATGGGGAATTCTCAACAAGTCTCCTGCGGAGCAGAAAT 687
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 CCAGTGTGCAGTGGTTGGTAATGGGGAATTCTCAACAAGTCTCCTGCGGAGCAGAAAT 240

Qy  688 TGATAAATCTGACTTCGTCTTCAGGTGTAACCTCCCCCCAATCACAGGGAGCGCTAGTAA 747
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 TGATAAATCTGACTTCGTCTTCAGGTGTAACCTCCCCCCAATCACAGGGAGCGCTAGTAA 300

Qy  748 AGATGTTGGAAGCAAAACAAATCTGTGACTGTCAATCCAGCATTATAACCCTGAAGTA 807
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 AGATGTTGGAAGCAAAACAAATCTGTGACTGTCAATCCAGCATTATAACCCTGAAGTA 360

Qy  808 CCAGAATTTGAAGGAGGAAGCACAGTTTTTGGAGACATCTCCACCTATGGAGATGC 867
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 CCAGAATTTGAAGGAGGAAGCACAGTTTTTGGAGACATCTCCACCTATGGAGATGC 420

Qy  868 ATTCCCTCCTGCCAGCATTTTCCTATCGGGCCAACACAGGCATCTCTTTTAAAGTCTA 927
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 ATTCCCTCCTGCCAGCATTTTCCTATCGGGCCAACACAGGCATCTCTTTTAAAGTCTA 480

Qy  928 CCAAACACTCAAAGAGTCAAAAATGAGGCAAAAGGTTCTCTTCTTCCATCCCAGGTACCT 987
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 CCAAACACTCAAAGAGTCAAAAATGAGGCAAAAGGTTCTCTTCTTCCATCCCAGGTACCT 540

Qy  988 GAGACACCTCGCTCTTTTCTGGAGAACTAAAGGGGTGCATACCGCTTGTCCACAGG 1047
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 GAGACACCTCGCTCTTTTCTGGAGAACTAAAGGGGTGCATACCGCTTGTCCACAGG 600

Qy 1048 CTTGATGATTGCAAGTGTCGCTGTGGAACTGTGTGAAAACGTGAAGCTTCACGGATTCTG 1107
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 CTTGATGATTGCAAGTGTCGCTGTGGAACTGTGTGAAAACGTGAAGCTTCACGGATTCTG 660

Qy 1108 GCCTTTCTCTAAGACTATCGAAGACAACCCCACTCAGTCACCTACTACTGATGATAACATGTT 1167
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661 GCCTTTCTCTAAGACTATCGAAGACAACCCCACTCAGTCACCTACTACTGATGATAACATGTT 720

Qy 1168 ACCTAAGCATGGTTCCACCAGATGCCTAAAGAATACAGCCAAAATGTCTCCAGCTCCATAT 1227
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721 ACCTAAGCATGGTTCCACCAGATGCCTAAAGAATACAGCCAAAATGTCTCCAGCTCCATAT 780

Qy 1228 GAGAGGAATCCTCAAACTGCAAATTCAGCAAATGTGAAACG 1267
        |||||||||||||||||||||||||||||||||||||||||
Db  781 GAGAGGAATCCTCANACTGCAAATTCAGCAAATGTGAAACG 820
```

RESULT 5
BG919404
LOCUS   BG919404   812 bp   mRNA   linear   EST 05-JUN-2001
DEFINITION 6028182955F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:4947286 5',
            mRNA sequence.
ACCESSION   BG919404
VERSION   BG919404.1 GI:14299880
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey Green M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10896 row: n column: 23
            High quality sequence stop: 795.
FEATURES   Location/Qualifiers
     source   1..812
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4947286"
                     /sex="female, virgin"
                     /tissue_type="infiltrating ductal carcinoma"
                     /dev_stage="5 months"
                     /lab_host="DH10"
                     /clone_lib="NCI_CGAP_Mam6"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
                     Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Jeffrey Green, M.D., NIH"
ORIGIN

```
Query Match        24.0%;   Score 760.4; DB 2; Length 812;
Best Local Similarity 98.6%;   Prd. No. 1.6e-185;
Matches 767;  Conservative 0;  Mismatches 11;  Indels 0;  Gaps 0;

Qy  164 CCTGCCCGTCCAGGCTGTTGA  GGAGGGAAGCAGAGAGGAGGAGGACACCAGTGGTACCTCAGCT 223
        |||  || |  |  |||||||  |||||||||||||||||||||||||||||||||||||||||
Db   33 CCGGAAAGATCCCGGGCTGTTGA GGAGGGAAGCAGAGAGGAGGAGGACACCAGTGGTACCTCAGCT 92

Qy  224 GCACTGAAGACACTCTGGAGCC GACAACCCCGGTACCACCGCACCAGGAACAGACAGACATAT 283
        |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db   93 GCACTGAAGACACTCTGGAGCC GACAACCCCGGTACCACCGCACCAGGAACAGACAGACATAT 152

Qy  284 CTGGATGAGAGAGACAACCCAAA AACAGAGAAATGCAATATAGCTTGAAC 343
        ||||||||||||||||||||||| ||||||||||||||||||||||||||
Db  153 CTGGATGAGAGAGACAACCCAAA AACAGAGAAATGCAATATAGCTTGAAC 212

Qy  344 TCTTTATCTAACAAAACGAGAC GTACTCTGAGGATGACTACCTCCAGACCATCACAAAC 403
        |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db  213 TCTTTATCTAACAAAACGAGAC GTACTCTGAGGATGACTACCTCCAGACCATCACAAAC 272

Qy  404 ATACAGAGATGCCCATGAACC GCAAGCAGAGAGAAGAATATGACAATTTTAGAGACAAAACTG 463
        ||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db  273 ATACAGAGATGCCCATGAACC GCAAGCAGAGAGAAGAATATGACAATTTTAGAGACAAAACTG 332

Qy  464 GCTTCCTGTGTTGCGATGCCATTC CAGACTTCGTGGTTTCCCAGAACAACACTCCAGTGGGG 523
        |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db  333 GCTTCCTGTGTTGCGATGCCATTC CAGACTTCGTGGTTTCCCAGAACAACACTCCAGTGGGG 392

Qy  524 ACTAACATGAGCTACGAGGTGG AAGCAAGAAACACATCCCATTCGAGAGAACATTTC 583
        |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db  393 ACTAACATGAGCTACGAGGTGG AAGCAAGAAACACATCCCATTCGAGAGAACATTTC 452

Qy  584 CACATGTTCCAGTGTCGCAGC TTTTGTGGACTATCCCTATAACCAGTGTGCAGTGGTT 643
        ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db  453 CACATGTTCCAGTGTCGCAGC TTTTGTGGACTATCCCTATAACCAGTGTGCAGTGGTT 512

Qy  644 GGTAATGGGGGAATTCTCAACA GTCTCCTGCGGAGCAGAAATTGATAAATCTGACTTC 703
```

```
Db          GGTAATGGGGAATTCTCAACAAGTCTCTGCGGAGCAGAAATTGATAAATCTGACTTC 572
Qy      704 GTCTTCAGTGTGTAACCTCCCCCCAATCACAGCAGGAGGGCTAGTAGAGATGTTGGAAGCAAA 763
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db      573 GTCTTCAGTGTGTAACCTCCCCCCAATCACAGCAGGAGCGCTAGTAGAGATGTTCGGAAGCAAA 632
Qy      764 ACAAATCTTGTGACTGTCAATCCCAGCATTATAACCCTGAAGTACCAGAATTTGAAGGAG 823
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      633 ACAAATCTTGTGACTGTCAATCCCAGCATTATAACCCTGAAGTACCAGAATTTGAAGGAG 692
Qy      824 AAGAAAGCACAGTTTTGGAGGACATCTCCACCTATGAGAGATGCATTCCTCCCTCCTGCCA 883
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      693 AAGAAAGCACAGTTTTCGGAGGACATCTCCACCTATGAGAGATGCATTCCTCCCTCCTGCCA 752
Qy      884 GCATTTTCCTATCCGGGCCAACACAGGCATCTCTTTTAAAGTCTACCAAACACTCAAAG 941
            ||||||||||||||||||||||||||||||||||||||||||||||||| ||| |||
Db      753 GCATTTTCCTATCGGGCCAACACAGGCATCTCTTTTAAAGTCTATCAAACAATCCAAG 810
```

RESULT 6
BI695458
LOCUS   BI695458   780 bp   mRNA   linear   EST 18-SEP-2001
DEFINITION 603346462F1 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:5374215 5',
            mRNA sequence.
ACCESSION   BI695458
VERSION   BI695458.1 GI:15658087
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 780)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11952 row: k column: 16
            High quality sequence stop: 780.
FEATURES   Location/Qualifiers
     source   1..780
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N-3"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5374215"
                     /tissue_type="tumor, biopsy sample"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Mam2"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
                     Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
ORIGIN

Query Match   23.6%; Score 748.4; DB 2; Length 780;
Best Local Similarity   99.9%; Pred. No. 2.1e-182;
Matches 749; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

```
Qy      845 GACATCTCCACCTATGAGAGATGCATTCCTCCCTCCTGCCAGCATTTTCCTATCCGGGCCAAC 904
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 GACATCTCCACCTATGAGAGATGCATTCCTCCCTCCTGCCAGCATTTTCCTATCGGGCCAAC 60
Qy      905 ACAGGCATCTCTTTTAAAGTCTACCAAACACTCAAAGTCTACCAAACACTCAAAAATGAGGCAGCAAAAGGTT 964
```

```
Db     659 CACAAGCATCATCATGCCAGTTTGAACACAGTGCTGAAAGGATCATAGACAGGGTGGT 718

Qy    2214 TAAATCTGATCCCAGTAGAATA 2236

Db     719 TAAATCTGATCCCAGTAGAATA 741
```

```
RESULT 8
BB486599
LOCUS       BB486599                 670 bp    mRNA    linear   EST 25-OCT-2001
DEFINITION  BB486599 RIKEN full-length enriched, 13 days embryo lung Mus
            musculus cDNA clone D430037E10 3', mRNA sequence.
ACCESSION   BB486599
VERSION     BB486599.2  GI:16441407
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 670)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T.,
            Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J.,
            Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K.,
            Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tanaka,A., Takahashi,F.,
            Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al 2001)
  JOURNAL   Unpublished (2001)
COMMENT     On Jul 23, 2000 this sequence version replaced gi:9404208.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center (GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
            wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Shikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Kawai,J., Ozazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
            and Hayashizaki,Y.
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
            Sugahara,Y. and Hayashizaki,Y.
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
            Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I.,
            Aizawa,K., Fukuda,S., Hara,A. Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
            Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mmm. Genome. 12, 673-677 (2001)
            Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
FEATURES             Location/Qualifiers
     source          1..670
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
```

```
http://image.llnl.gov
Plate: LLAM11867  row: 1  column: 08
High quality sequence start: 5
High quality sequence stop: 749.
FEATURES             Location/Qualifiers
     source          1..753
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="129,C57BL/6J,FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5341591"
                     /tissue_type="tumor, gross tissue"
                     /dev_stage="10 months"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Mam3"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigators
                     providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
                     Reference for transgenic model: Xu et al., Nature Genetics
                     22, 37-43 (1999)."
ORIGIN

Query Match       22.7%;  Score 718;  DB 2;  Length 753;
Best Local Similarity  99.6%;  Pred. No. 1.6e-174;
Matches 740;  Conservative   0;  Mismatches   0;  Indels  1;  Gaps   2;

Qy  1494 CTTGGGATAAGTAAAATGAGATAATTTTTCAAATCATCATAGAATTTGATTTGAGCCAGGG 1553
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db     1 CTTGGGATAAGTAAAATGAGATAATTTTTCAAATCATCATAGAATTTGATTTGAGCCA-GG 59

Qy  1554 TCTCTCAGAATGCTTCCTTGTTCCTATCCATGATGAGCCATTCCCACCTTTATCAGAGTGG 1613
         |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db    60 TCTCTCAGAATGCTTCCTTG-TCCTATCCATGATGAGCCATTCCCACCTTTATCAGAGTGG 118

Qy  1614 TAATGAAACTGTGCAATGTGCCAAAGACCCTTTCTGAAGAGAATGTCTGAATCATGCGC 1673
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   119 TAATGAAACTGTGCAATGTGCCAAAGACCCTTTCTGAAGAGAATGTCTGAATCATGCGC 178

Qy  1674 CGAGTTTTACACCAGCTCCTCCTTTATAAATAATCCTTCCCATTCTCCCCCTAGTA 1733
         ||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   179 CGAGTTTTACACAGCTCTTCCTTTATAAATAATCCTTCCCATTCTCCCCTCTAGTA 238

Qy  1734 GAGTACAGAAACAAAATACCCTTGATGATTCAGGAAGAAAGTCTTTTTTACTTAGCAAT 1793
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   239 GAGTACAGAAACAAAATACCCTTGATGATTCAGGAAGAAAGTCTTTTTTACTTAGCAAT 298

Qy  1794 GTGCCTGCTTCTGATTCAGTTCGCTTGTGACATTAAGCTGGGTTTGGGTTTGGTTGGAT 1853
         |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   299 GTGCCTGCTTCTGATTCAGTTCGCTTGTGACATTAAGCTGGGTTTGGGTTTNNGTTGGAT 358

Qy  1854 TTTGGGGCGCTTCTTCACTTCTTTTGTCTATATTTTCCTTACCTTATCAGTTTGTATTCG 1913
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   359 TTTGGGGCGCTTCTTCACTTCTTTTGTCTATATTTTCCTTACCTTATCAGTTTGTATTCG 418

Qy  1914 AGCTTCCTGCTTGGGATTCGCAATTCTCTCTCCCACTGACAGGATCAACTCAATGACA 1973
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   419 AGCTTCCTGCTTGGGATTCGCAATTCTCTCTCCCACTGACAGGATCAACTCAATGACA 478

Qy  1974 TAAAGTAGTTCAAACATCCATTGCTTCTCCACATGTTTTATCCATAAAGTACTCATCTGA 2033
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   479 TAAAGTAGTTCAAACATCCATTGCTTCTCCACATGTTTTATCCATAAAGTACTCATCTGA 538

Qy  2034 TTTTATTTAAAATAGTAGAACATCCTACTTGATATCAGACCCGAGGACCATCCTCCATTGGA 2093
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   539 TTTTATTTAAAATAGTAGAACATCCTACTTGATATCAGACCCGAGGACCATCCTCCATTGGA 598

Qy  2094 GAATATGAAGATATTGTCACTGGCAGAAAAGCCAGGTGTGTGCCATTAATTGATAAGATAC 2153
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   599 GAATATGAAGATATTGTCACTGGCAGAAAAGCCAGGTGTGTGCCATTAATTGATAAGATAC 658

Qy  2154 CACAAGCATCATCATGCCAGTTATGAACACAGTGCTGAAAGGATCATAGACAGGGTGGT 2213
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
                                              Db          661 TCATGCTATT 670
        /clone="D430037E10"
        /tissue_type="lung"                   RESULT 9
        /dev_stage="13 days embryo"           AZ315725/c
        /lab_host="DH10B"                     LOCUS       AZ315725          640 bp    DNA       linear   GSS 29-SEP-2000
        /clone_lib="RIKEN full-length enriched, 13 days embryo   DEFINITION  1M0032M24R Mouse 10kb plasmid UUGCIM library Mus musculus genomic
        lung"                                             clone UUGCIM0032M24 R, genomic survey sequence.
        /note="Site 1: SalI; Site_2: BamHI; cDNA library was    ACCESSION   AZ315725
        prepared and sequenced in Mouse Genome Encyclopedia     VERSION     AZ315725.1  GI:10362673
        Project of Genome Exploration Research Group in Riken    KEYWORDS    GSS.
        Genomic Sciences Center and Genome Science Laboratory in SOURCE      Mus musculus (house mouse)
        RIKEN. Division of Experimental Animal Research in Riken   ORGANISM  Mus musculus
        contributed to prepare mouse tissues. 1st strand cDNA was           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
        primed with a primer (5'                                            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
        GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTVN 3'], cDNA was            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
        prepared by using trehalose thermo-activated reverse     REFERENCE   1 (bases 1 to 640)
        transcriptase and subsequently enriched for full-length by AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
        cap-trapper. Second strand cDNA was prepared with the               Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
        primer adapter of sequence (5'                                      Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von
        GAGAGAGAGATTCGAGTTAATTAAATTAAATAATCCCCCCCCCCCC 3']. cDNA            Niederhausern,A. and Wright,D.,Weiss,R.
        was cleaved with BamHI and XhoI. Vector: a modified      TITLE       Mouse whole genome scaffolding with paired end reads from 10kb
        pBluescript KS(+) after bulk excision from Lambda FLC I."            plasmid inserts
ORIGIN                                        JOURNAL     Unpublished (2000)
                                              COMMENT     Contact: Robert B. Weiss
Query Match      21.1%;     Score 666.8; DB 7; Length 670;            University of Utah Genome Center
Best Local Similarity   99.7%;  Pred. No. 3e-161;                    University of Utah
Matches  668;  Conservative  0;  Mismatches  2;  Indels  0;  Gaps  0;   Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
                                                                     84112, USA
Qy  2483 GAGTAATTCAGAGCCAATCACATCACAAGGATGGACACACTAACTCATCACTTCAGGGG 2542   Tel: 801 585 5606
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||   Fax: 801 585 7177
Db     1 GAGTAATTCAGAGCCAATCACATCACAAGGATGGACACACTAACTCATCACTTCAGGGG 60    Email: ddunn@genetics.utah.edu
                                                                     Insert Length: 10000   Std Error: 0.00
Qy  2543 GAGATGAATGCTTCATGAGAAATTACACTCATAAGCTAAGCATCAGTTTGAGTAAAAT 2602    Plate: 0032   row: M  column: 24
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||   Seq primer: CACACAGGAAACAGCTATGACC
Db    61 GAGATGAATGCTTCATGAGAAATTACACTCATAAGCTAAGCATCAGTTTGAGTAAAAT 120    Class: plasmid ends
                                                                     High quality sequence stop: 640.
Qy  2603 TTGAGTAGATGTTAAATATGAACATTTTATACCTTCTTACTAAGTCCCACCGACACCTTT 2662   Location/Qualifiers
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||   FEATURES
Db   121 TTGAGTAGATGTTAAATATGAACATTTTATACCTTCTTACTAAGTCCCACCGACACCTTT 180    source          1..640
                                                                                      /organism="Mus musculus"
Qy  2663 TAATGTAAGCACATTTATTTATTAAGTTACTTGACATTAAATGCTTATGTCTGTATATTC 2722                  /mol_type="genomic DNA"
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||                  /strain="C57BL/6J"
Db   181 TAATGTAAGCACATTTATTTATTAAGTTACTTGACATTAAATGCTTATGTCTGTATATTC 240                  /db_xref="taxon:10090"
                                                                                      /clone="UUGCIM0032M24"
Qy  2723 TGTTCATCCATCGATTTCCCAAAAAGTAAGAGCATAGGAGATGAGGCCTACATGCCAAG 2782                  /sex="Male"
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||                  /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
Db   241 TGTTCATCCATCGATTTCCCAAAAAGTAAGAGCATAGGAGATGAGGCCTACATGCCAAG 300                  /clone_lib="Mouse 10kb plasmid UUGCIM library"
                                                                                      /note=Vector: pWD42nv; Purified genomic DNA from M.
Qy  2783 AAAACTATAAATTTTACTCTTTAATTCTTACTTGAGCCAGCTTGTTGTTTATCAAGTGCT 2842              musculus C57BL/6J (male) was obtained from the Jackson
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||                  Laboratory Mouse DNA Resource
Db   301 AAAACTATAAATTTTACTCTTTAATTCTTACTTGAGCCAGCTTGTTGTTTATCAAGTGCT 360                  (http://www.jax.org/resources/documents/dnares/). The DNA
                                                                                      was hydrodynamically sheared by repeated passage through a
Qy  2843 TTTTTGAAGAGCAGCACCCTGGAATTCTCATTCTGATACAGTACAGTGTCACCTTGTATTA 2902             0.005 inch orifice at constant velocity. The sheared DNA
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||                  was blunt end-repaired with T4 DNA polymerase and T4
Db   361 TTTTTGAAGAGCAGCACCCTGGAATTCTCATTCTGATACAGTACAGTGTCACCTTGTATTA 420             polynucleotide kinase. Adaptor oligonucleotides were
                                                                                      ligated to the blunt ends in high molar excess. The
Qy  2903 ACATTTGTAAATGTTGTTTCAAGTTTACATCTCTTCATTCTTTTATAGCAAATCAAACGT 2962              adaptored DNA was purified and size-selected for a 9.5 to
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||                  10.5 kb range using preparative agarose gel
Db   421 ACATTTGTAAATGTTGTTTCAAGTTTACATCTCTTCATTCTTTTATAGCAAATCAAACGT 480              electrophoresis. Vector DNA was prepared from a derivative
                                                                                      of pWD42 (gi|47|2114|gb|AF129072.1), a copy-number
Qy  2963 ATTAGCTTCAGAAATTATCAGAAGTTCATATATAAATATTTGCAAAGGGTAAAAGGCT 3022                inducible derivative of plasmid R1. The vector was ligated
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||                  with adaptors complementary to the insert adaptors and
Db   481 ATTAGCTTCAGAAATTATCAGAAGTTCATATATAAATATTTGCAAAGGGTAAAAGGCT 540              purified. The sheared, adaptored mouse DNA was annealed to
                                                                                      adaptored vector DNA, and transformed into
Qy  3023 TTTTTGTTAAATAAAATAAAATTTATTATTTTCTTCGTGATGAATAGAGGCTCTTTTATGC 3082          chemically-competent E. coli XL10-Gold (Stratagene) cells
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||                  and selected for ampicillin resistance."
Db   541 TTTTTGTTAAATAAAATAAAATTTATTATTTTCTTCGTGATGAATAGAGGCTCTTTTATGC 600   ORIGIN
                                                                     Query Match       19.9%;    Score 631;  DB 11;  Length 640;
Qy  3083 TGCTGCTAATGAACCTAATTAGCTTAAATTATCTCCTAGCAACATTGGTCACGTTTCAA 3142   Best Local Similarity  100.0%;  Pred. No. 5.9e-152;
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||   Matches  631;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;
Db   601 TGCTGCTAATGAACCTAATTAGCTTAAATTATCTCCTAGCAACATTGGTCACGTTTCAA 660
                                                                     Qy   1283 GAAGGAGAATAATTCAGGAGGCGGAGTGGAATGGTCACAGCATCTCCAAAAGCCAATA 1342
Qy  3143 TCATGCTATT 3152
         ||||||||||
```

```
                                    cDNA library was prepared and sequenced in Mouse Genome
                                    Encyclopedia Project of Genome Exploration Research Group in Riken
                                    Genomic Sciences Center and Genome Science Laboratory in RIKEN.
                                    Division of Experimental Animal Research in Riken contributed to
                                    prepare mouse tissues.
                                    Please visit our web site for further details.
                                    URL:http://genome.gsc.riken.jp/.
FEATURES             Location/Qualifiers
     source          1..628
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="9530036B06"
                     /sex="male"
                     /tissue_type="urinary bladder"
                     /dev_stage="adult"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library,
                     C57BL/6J urinary bladder male adult"
ORIGIN

Query Match         19.7%;  Score 624.8;  DB 7;  Length 628;
Best Local Similarity   99.7%;  Pred. No. 2.4e-150;
Matches   626;  Conservative     0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1361 GCATGAATTACAAAGGGCTCTCCACTTGTCTAGACCAAAGCCACCGCCCCCACTCAC 1420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 GCATGAATTACAAAGGGCTCTCCACTTGTTTAGACCAAAGCCACCGCCCCCAATCAC 60

Qy      1421 TTTGCAGCTCCACGAGTCACTCATTCCACCTTCAAGTTCTTCTCTGAGAATAGAGA 1480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 TTTGCAGCTCCACGAGTCACTCATTCCACCTTCAAGTTCTTCTCTGAGAATAGAGA 120

Qy      1481 CCAAAACATCAGACTTGGATAATAAAATGAGATAATTTTCAAATCATCATAGAATTTG 1540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 CCAAAACATCAGACTTGGATAATAAAATGAGATAATTTTCAAATCATCATAGAATTTG 180

Qy      1541 ATTTGAGCCAGGTCTCTCAGATGCTTCCTTGTCCTATCCATCGATAGCCATTCCCACC 1600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 ATTTGAGCCAGGTCTCTCAGATGCTTCCTTGTCCTATCCATCGATAGCCATTCCCACC 240

Qy      1601 TTTATCAGAGTGGTAATGAAACGTGCAATTGTGCCAAAGACCCTTTCGAAGAGAATGT 1660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 TTTATCAGAGTGGTAATGAAACGTGCAATTGTGCCAAAGACCCTTTCGAAGAGAATGT 300

Qy      1661 CTGAATCATGCGCCGAGTTTTCACACAGCTCTTCCTTTATAAATAAATCCTTCCCATT 1720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 CTGAATCATGCGCCGAGTTTTCACACAGCTCTTCCTTTATAAATAAATCCTTCCCATT 360

Qy      1721 CTCCCTCCTAGTAGAGTACAGAACAAAATACCCTGATGATTCAGGAGGAAAAGTCTTT 1780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 CTCCCTCCTAGTAGAGTACAGAACAAAATACCCTGATGATTCAGGAGGAAAAGTCTTT 420

Qy      1781 TTTACTTAGCAATGTGCCTGCTCTGATTCAGTTCGCTTGTGACATTAAGCTGGGTTGGG 1840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 TTTACTTAGCAATGTGCCTGCTCTGATTCAGTTCGCTTGTGACATTAAGCTGGGTTGGG 480

Qy      1841 GTTTTGGTTGGATTTGGGGCGTTCTTCACTTCTTTTGTCTATATTTTCCTTACCTTTAT 1900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 GTTTTGGTTGGATTTGGGGCGTTCTTCACTTCTTTTGTCTATATTTTCCTTACCTTTAT 540

Qy      1901 CAGTTTGTATTCGAGCTTCCTGTTTGGGATTCGCAATTCCTCTCCCCACTGACAGGAT 1960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 CAGTTTGTATTCGAGCTTCCTGTTTGGGATTCGCAATTCCTCTCCCCACTGACAGGAT 600

Qy      1961 CAACTCAATGACATAAAGTAGTCAAAC 1988
             ||||||||||||||||||||||||||
Db       601 CAACTCAATGACATAAAGTAGTCAAAC 628
```

```
Db       640 GAAGGAGAATAATTTCAGGAGGTGGATGTGTCACAGCATCTCCAAAAAGCCAATA 581
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1343 GAAGGAGGCACAGAGAAAGCATGAATTACAAAGGCGCTCTCCACTTGTCTAGACCAAAG 1402
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       580 GAAGGAGGCACAGAGAAAGCATGAATTACAAAGGCGCTCTCCCACTTGTCTAGACCAAAG 521

Qy      1403 CCACCCGCCCCCACTCACTTTGCAGCTCCACGAGTCACTCATTCTCACCTTCAACGTTC 1462
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       520 CCACCCGCCCCCACTCACTTTGCAGCCTCCACGAGTCACTCATTCTCACCTTCAACGTTC 461

Qy      1463 TTTCTCTGAGAATAGAGACCAAAACATCAGACTTGGATAAGTAAAATGAGATAATTTTC 1522
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       460 TTTCTCTGAGAATAGAGACCAAAACATCAGACTTGGATAAGTAAAATGAGATAATTTTC 401

Qy      1523 AAATCATCATAGAATTTGATTTGAGCCAGGGTCTCTCAGAATGCTTCCTTGTTCCTATCC 1582
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       400 AAATCATCATAGAATTTGATTTGAGCCAGGGTCTCTCAGAATGCTTCCTTGTTCCTATCC 341

Qy      1583 ATGATAGCCATTCCCACCTTTATCAGAGTGGTAATGAAACTGTGCAATTGTGCCAAAGAC 1642
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       340 ATGATAGCCATTCCCACCTTTATCAGAGTGGTAATGAAACTGTGCAATTGTGCCAAAGAC 281

Qy      1643 CCTTTCGAAGAGAATGTCTGAATCATGCGCCGAGTTTTTACACACAGCTCTTCCTTTAT 1702
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       280 CCTTTCGAAGAGAATGTCTGAATCATGCGCCGAGTTTTTACACACAGCTCTTCCTTTAT 221

Qy      1703 AAATAAATCCTTCCCATTCTCCCTCCTAGTAGAGTACAGAACAAAATACCCTGATGAT 1762
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       220 AAATAAATCCTTCCCATTCTCCCTCCTAGTAGAGTACAGAACAAAATACCCTGATGAT 161

Qy      1763 TCAGGAAGAAAAGTCTTTTTTACTTAGCAATGTGCCTGCTCTGATTCAGTTCGCTTGTG 1822
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       160 TCAGGAAGAAAAGTCTTTTTTACTTAGCAATGTGCCTGCTCTGATTCAGTTCGCTTGTG 101

Qy      1823 ACATTAAGCTGGGTTGGGGTTTTGGTTGGATTTGGGGCGTTTCTTCACTTCTTTTGTCTA 1882
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       100 ACATTAAGCTGGGTTGGGGTTTTGGTTGGATTTGGGGCGTTTCTTCACTTCTTTTGTCTA 41

Qy      1883 TATTTTCCTTACCTTTATCAGTTTGTATTCG 1913
             ||||||||||||||||||||||||||||||
Db        40 TATTTTCCTTACCTTTATCAGTTTGTATTCG 10
```

```
RESULT 10
BB112539
LOCUS       BB112539                 628 bp    mRNA    linear    EST 08-NOV-2005
DEFINITION  BB112539 RIKEN full-length enriched mouse cDNA library, C57BL/6J
            urinary bladder male adult Mus musculus cDNA clone 9530036B06 3',
            mRNA sequence.
ACCESSION   BB112539
VERSION     BB112539.2  GI:15408053
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 628)
  AUTHORS   The FANTOM Consortium and RIKEN Genome Exploration Research Group
            and Genome Science Group (Genome Network Project Core Group).
  TITLE     The transcriptional landscape of the mammalian genome
  JOURNAL   Science 309 (5740), 1559-1563 (2005)
   PUBMED   16141072
COMMENT     On Jun 27, 2000 this sequence version replaced gi:8765107.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
```

```
RESULT 11
BB627781
LOCUS       BB627781                 625 bp    mRNA    linear    EST 08-NOV-2005
```

```
DEFINITION   BB627781 RIKEN full-length enriched mouse cDNA library, C57BL/6J
             urinary bladder male adult Mus musculus cDNA clone 9530036E06 5',
             mRNA sequence.
ACCESSION    BB627781
VERSION      BB627781.1  GI:1539924
KEYWORDS     EST.
SOURCE       Mus musculus (house mouse)
ORGANISM     Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
             Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 625)
AUTHORS      The FANTOM Consortium and RIKEN Genome Exploration Research Group
             and Genome Science Group (Genome Network Project Core Group).
TITLE        The transcriptional landscape of the mammalian genome
JOURNAL      Science 309 (5740), 1559-1563 (2005)
PUBMED       1614 1072
COMMENT      Contact: Yoshihide Hayashizaki
             Laboratory for Genome Exploration Research Group, RIKEN Genomic
             Sciences Center (GSC), Yokohama Institute
             The Institute of Physical and Chemical Research (RIKEN)
             1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
             Tel: 81-45-503-9222
             Fax: 81-45-503-9216
             Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
             cDNA library was prepared and sequenced in Mouse Genome
             Encyclopedia Project of Genome Exploration Research Group in Riken
             Genomic Sciences Center and Genome Science Laboratory in RIKEN.
             Division of Experimental Animal Research in Riken contributed to
             prepare mouse tissues.
             Please visit our web site for further details.
             URL:http://genome.gsc.riken.jp/.
FEATURES     Location/Qualifiers
     source  1..625
             /organism="Mus musculus"
             /mol_type="mRNA"
             /strain="C57BL/6J"
             /db_xref="taxon:10090"
             /clone="9530036E06"
             /sex="male"
             /tissue_type="urinary bladder"
             /dev_stage="adult"
             /clone_lib="RIKEN full-length enriched mouse cDNA library,
             C57BL/6J urinary bladder male adult"
ORIGIN

Query Match       19.1%;  Score 604;  DB 7;  Length 625;
Best Local Similarity  100.0%;  Pred. No. 6e-145;
Matches   604;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy     1 CGGAGCGGCGGCGAGTCGGTGCCGCCCCGGCGCTTGCGCTTCGCCCGGCAGCTTGGCGGCGAGG   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    21 CGGAGCGGCGGCGAGTCGGTGCCGCCCCGGCGCTTGCGCTTCGCCCGGCAGCTTGGCGGCGAGG   80

Qy    61 ACGCCCGTGGCTCAGGATGAGATCGGGGGCACGCTGTTCGCCCTCATAGGCAGCCTGAT  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    81 ACGCCCGTGGCTCAGGATGAGATCGGGGGCACGCTGTTCGCCCTCATAGGCAGCCTGAT  140

Qy   121 GCTGCTGCTCCTCCTCGGGTATGCTCTGGTGCCCAGCCGACGGCCGCTGCGCCCGCTTCCAGGCT  180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   141 GCTGCTGCTCCTCCTCGGGTATGCTCTGGTGCCCAGCCGACGGCCGCTGCGCCCGCTTCCAGGCT  200

Qy   181 GTTGATGAGGGAAGCAGAGAGGAGACACCCAGTGTACCTCAGCTGCACTGAAGACACTCTG  240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   201 GTTGATGAGGGAAGCAGAGAGGAGACACCCAGTGTACCTCAGCTGCACTGAAGACACTCTG  260

Qy   241 GAGCCCACAACCCCGGTACCACGGACACCACCAGGAACAGCACATATCTGGATGAGAGACAC  300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   261 GAGCCCGACAACCCCGGTACCACGGACACCACCAGGAACAGCACATATCTGGATGAGAGACAAC  320

Qy   301 CCAAATAACAGAGAAATGCAAAGATCTGCAATATAGCTTGAACTCTTTATCTTAACAAAAC  360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   321 CCAAATAACAGAGAAATGCAAAGATCTGCAATATAGCTTGAACTCTTTATCTTAACAAAAC  380
```

```
Qy   361 GAGACGGTACTCTGAGGATGAC TACCTTCCAGACATCACAAACATACAGAGATGCCCATG  420
         ||||||||||||||||||||||     ||||||||||||||||||||||||||||||||||
Db   381 GAGACGGTACTCTGAGGATGAC TACCTTCCAGACATCACAAACATACAGAGATGCCCATG  440

Qy   421 GAACCCGGCAAGCAGAGAAT ATACAATTTTAGAGCAAAACTGGCTTCCTCGTTGGCGATGC  480
         ||||||||||||||||||||     ||||||||||||||||||||||||||||||||||||
Db   441 GAACCCGGCAAGCAGAGAAT ATACAATTTTAGAGCAAAACTGGCTTCCTCGTTGGCGATGC  500

Qy   481 CATTCAAGACTTCGTCGTTCC CAGAACAACACTCCAGTGGGGACTAACATGAGCTACGA  540
         |||||||||||||||||||||     |||||||||||||||||||||||||||||||||||
Db   501 CATTCAAGACTTCGTCGTTCC CAGAACAACAACACTCCAGTGGGGACTAACATGAGCTACGA  560

Qy   541 GGTGGAAAGCAAGAAACACATC CCATTCGAGAGAACATTTTCCACATGTTTCCAGTGTC  600
         ||||||||||||||||||||||     ||||||||||||||||||||||||||||||||||
Db   561 GGTGGAAAGCAAGAAACACATC CCATTCGAGAGAACATTTTTCCACATGTTTCCAGTGTC  620

Qy   601 GCAG  604
         ||||
Db   621 GCAG  624
```

```
RESULT 12
BB633149/c
LOCUS        BB633149  631 bp  mRNA  linear  EST 25-AUG-2000
DEFINITION   uv73e09.y1 Soares mouse 3NDMS Mus musculus cDNA clone IMAGE:3412840
             5' similar to TR:P97713  P97713 GD3 SYNTHASE ;, mRNA sequence.
ACCESSION    BB633149
VERSION      BB633149.1  GI:9915837
KEYWORDS     EST.
SOURCE       Mus musculus (house mouse)
ORGANISM     Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
             Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 631)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE        National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             MGI:1088652
             Possible reversed clone: similarity on wrong strand
             Seq primer: -40RP from Gibco
             High quality sequence stop: 467.
FEATURES     Location/Qualifiers
     source  1..631
             /organism="Mus musculus"
             /mol_type="mRNA"
             /strain="C57BL/6J"
             /db_xref="taxon:10090"
             /clone="IMAGE:3412840"
             /sex="male"
             /tissue_type="Spleen"
             /dev_stage="4 weeks"
             /lab_host="DH10B"
             /clone_lib="Soares mouse 3NDMS"
             /note="Vector: T7T3D-PacI; Site 1: Not I; Site 2: Eco RI;
             1st strand cDNA was primed with a Not I - oligo(dT) primer
             [5'
             TGTTACCAATCTGAATGGAGCGGCGCCGCGTGTTTTTTTTTTTTTTTTTTTTT
             3'], double-stranded cDNA was ligated to Eco RI adaptors
             (Pharmacia), digested with Not I and cloned into the Not I
             and Eco RI site of the modified pT7T3 vector. RNA
             provided by Dr. Bertrand Jordan. Library went through
             three rounds of normalization, and was constructed by
             Bento Soares and M.Fatima Bonaldo."
ORIGIN

Query Match       18.8%;  Score 596.4;  DB 7;  Length 631;
```

```
Best Local Similarity  96.5%;  Pred. No. 5.7e-143;
Matches  609;  Conservative 0;  Mismatches 22;  Indels 0;  Gaps 0;

Qy   755 GGAAGCAAACAAATCTTGTGACTGTCAATCCCAGCAATATAACCCTGGAAGTACCAGAAT 814
         ||||||||||  ||||||||||||||||||  ||||||| ||||||| ||||||||||||
Db   631 GGAAGCAAACCCAATCTTGTGACTGTCAATCCAGCAATTATAACCCTGAAGTACCAGAAT 572

Qy   815 TTGAAGGAGGAAGAAGCACAGTTTTGGAGGAGCACATCCCACCTATGGAGATGCATTCCTC 874
         |||||||||||||||||||||||||||||||||||||| ||||||||| |||||||||||
Db   571 TTGAAGGAGGAAGAAGCACAGTTTTGGAGGAGCACATCTCCACCTATGNAGATGCATTCCTC 512

Qy   875 CTCCTGCCAGCATTTCCTATCGGGCCAACACAGGCATCTCTTTAAAGTCTACCAAACA 934
         |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   511 CTCCTGCCAGCATTTCCTATCGGGCCAACACAGGCATCTCTTTTAAAGTCTACCAAACA 452

Qy   935 CTCAAAGAGTCAAAAATGAGGCAAAAGGTTCTCTTCTTCCATCCCAGGTACCTGAGACAC 994
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   451 CTCAAAGAGTCAAAAATGAGGCAAAAGGTTCTCTTCTTCCATCCCAGGTACCTGAGACAC 392

Qy   995 CTCGCTCTTTTCTGGAGAACTAAAGGGGTGACTGCATACCGCTTGTCCACAGGCTTGATG 1054
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   391 CTCGCTCTTTTCTGGAGAACTAAAGGGGTGACTGCATACCGCTTGTCCACAGGCTTGATG 332

Qy  1055 ATTGCAAGTGTCCGCTGTGGAACTGTGTGAAAACGTGAAGCTCACGGATTCTGGCCTTTC 1114
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   331 ATTGCAAGTGTCCGCTGTGGAACTGTGTGAAAACGTGAAGCTCACGGATTCTGGCCTTTC 272

Qy  1115 TCTAAGACTATCGAAGACACCCCACTCCAGTCACCACTACTATGATAACATGTTACCTAAG 1174
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   271 TCTAAGACTATCGAAGACACCCCACTCCAGTCACCACTACTATGATAACATGTTACCTAAG 212

Qy  1175 CATGGTTTCCACCAGATGCCTAAAGAATTACAGCCAAATGCTCCAGGTCCCATATGAGAGGA 1234
         |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db   211 CATGGTTTCCACCAGATGCCTAAAGAATACAGCCAAATGCTCCAGGTCCCATATGAGAGGA 152

Qy  1235 ATCCTCAAACTGCAATTCAGCAAATGTGAAACGGCTTAACGTTTCTTAGAGGAGAATAA 1294
         |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db   151 ATCCTCAAACTGCAATTCAGCAAATGTGAAACGGCTTAACGTTTCTTAGAGGAGAATAA 92

Qy  1295 TTTCAGGAGGTGAGTGGATGTGTCACAGCATCTCCAAAAGCCAATAGAAGAGGCCACA 1354
         |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    91 TTTCAGGAGGTGAGTGGATGTGTCACAGCATCTCCAAAAAGCCAATAGAAGAGGCCACA 32

Qy  1355 GAGAAAGCATGAATTACAAAAGGCGCTCTCCC 1385
         ||| | |  || |  ||||||||||||||||
Db    31 AAAAAAAAAAAAACAGGCGCGGCAGCTCCC 1
```

```
RESULT 13
BY752002
LOCUS       BY752002               663 bp    mRNA    linear   EST 17-DEC-2002
DEFINITION  BY752002 RIKEN full-length enriched, activated spleen Mus musculus
            cDNA clone F830206C17 5', mRNA sequence.
ACCESSION   BY752002
VERSION     BY752002.1  GI:27182667
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 663)
AUTHORS     Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
            Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
            Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
            Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
            Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
            Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
            Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
            Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
            Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
            Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
            Kurochkin,I.V., Lee,Y.,
            Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
            Numata,K., Okido,T., Pav   n,W.J., Pertea,G., Pesole,G.,
            Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
            Ravasi,T., Reed,J.C., Reid,D.J., Ring,B.Z., Ringwald,M.,
            Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
            Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
            Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
            Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
            Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
            Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
            Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
            Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
            Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
            Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
            Rogers,J., Birney,E. and Hayashizaki,Y.
TITLE       Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
JOURNAL     Nature 420, 563-573 (2002)
PUBMED      12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center (GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Adachi,J., Aizawa,K., Akmura,T., Arakawa,T., Carninci,P.,
            Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y.,
            Kondo,S., Konno,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M.,
            Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N.,
            Sano,H., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M.,
            Takeda,Y., Waki,K., Watanki,A., Muramatsu,M. and Hayashizaki,Y.
            Direct Submission
            Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mmm. Genome. 12, 673-677 (2001)
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (1), 1617-1630 (2000)
            RIKEN integrated sequence analysis (RISA) system-384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
            Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
            Trust/MRC building Addenbrookes Hospital Cambridge) whose
            assistance we gratefully acknowledge.
            Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1..663
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="NOD"
                     /db_xref="taxon:10090"
                     /clone="F830206C17"
                     /tissue_type="activated spleen"
                     /clone_lib="RIKEN full-length enriched, activated spleen"
ORIGIN

Query Match             18.2%;  Score 575.4;  DB 4;  Length 663;
Best Local Similarity   97.8%;  Prid No. 1.6e-137;
Matches  590;  Conservative 0;  Mismatches 12;  Indels 1;  Gaps 1;
```

```
Qy    1 CGGACGGGCGGAGTCGGTGCCGCCCCGGCGTGGCTTCGCCCCGGCAGCTTTGGCGGCGAGG   60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62 CGGAGCGGGCGGAGTCGGTGCCGCCCCGGCGTGGCTTCGCCCCGGCAGCTTTGGCGGCGAGG  121

Qy   61 ACGCCCGTGCTGCTCAGGATGAGATCGGGGGGCACGCTGTTCGCCCTCATAGGCAGCCTGAT  120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  122 ACGCCCGTGCTGCTCAGGATGAGATCGGGGGGCACGCTGTTCGCCCTCATAGGCAGCCTGAT  181

Qy  121 GCTGCTGCCTCCTCCGGTATGCTTCGGTGTGCCAGCGACGCGCCCGCTCCCGGCCTCCAGGCT  180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  182 GCTGCTGCTCCTCCGGTATGCTTCGGTGTGCCAGCGACGCGCCCGCTCCCGGCCTCCAGGCT  241

Qy  181 GTTGATGGAGGGAGCAGAGAGGAGACCACCAGTGGTACCTGCTGGTACTGCGACTGAAGACACTCTG  240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  242 GTTGATGGAGGGAGCAGAGAGGAGACCACCAGTGGTACCTGCTGGTACTGCGACTGAAGACACTCTG  301

Qy  241 GAGCCCGACAACCCCGGTACCAGGCACCAGGAACAGCACATATCTGGATGAGAAGACAAC  300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  302 GAGCCCGACAACCCCGGTACCACGCACCAGGAACAGCACATATCTGGATGAGAAGACAAC  361

Qy  301 CCAAATAACAGAGAAATGCAAAGATCTGCAATATAGCTTGAACTTTTTATCTAACAAAAC  360
        ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db  362 CCAAATAACAGAGAAATGCAAAGATCTGCAATATAGCTTGAACTTTTTATCTAACAAAAC  421

Qy  361 GAGACGGTACTCTGAGGATGACTACCTCCAGACCATCACAAACATACAGAGATGCCCATG  420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  422 GAGACGGTACTCTGAGGATGACTACCTCCAGACCATCACAAACATACAGAGATGCCCATG  481

Qy  421 GAACCGGCAAGCAGCAGAAGAATATGACAAATTTAGACAAGAAACTGGCTTCCTGTTGCGATGC  480
        |||| ||||||| |||||||||||||||||||||||||||| ||||||||||||||||||||
Db  482 GAACCGGNCAGCAGCAGAAGAATATGACAAATTNTAGACAAGAAACTGGCTTCCTGTTGCGATGC  541

Qy  481 CATTCAAGACTTCGTGGTTTCCCAGAACAACACTCCAGTGGGGACTAACAACATGAGCTACGA  540
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  542 CATTCAAGACTTCGTGGTTTCCCAGAAC-ACACTCCAGTGGGGACTAACAACATGAGCTACGA  600

Qy  541 GGTGGAAAGCAAGAAGACACATCCCCATTCGAGAGAAACATTTTCCACACATGTTTCCAGTGTC  600
        |||||||| |||| |||||||||||||||||||||||||||||||||| |||||||||||||
Db  601 GGTGGAAANGCANGAACACATCCCCATTCGAGAGAAACATTTTNCACACATGTTTNCAGTGTC  660

Qy  601 GCA  603
        |||
Db  661 ACA  663
```

RESULT 14
BI653246
BI653246
LOCUS   BI653246   852 bp   mRNA   linear   EST 12-SEP-2001
DEFINITION 603300931F1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5341615 5',
            mRNA sequence.
ACCESSION   BI653246
VERSION   BI653246.1   GI:15567482
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 852)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/,
            National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE   Unpublished (1999)
JOURNAL   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
COMMENT   Plate: LLAM11867 row: m column: 08
            High quality sequence start: 12

FEATURES   Location/Qualifiers
            source   1..852
                            /organism="Mus musculus"
                            /mol_type="mRNA"
                            /strain="129,C57BL/6J,FVB/N"
                            /db_xref="taxon:10090"
                            /clone="IMAGE:5341615"
                            /tissue_type="tumor, gross tissue"
                            /dev_stage="10 months"
                            /lab_host="DH10B"
                            /clone_lib="NCI_CGAP_Mam3"
                            /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
                            Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT;
                            Library constructed by Life Technologies. Investigators
                            providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
                            Reference for transgenic model: Xu et al., Nature Genetics
                            22, 37-43 (1999)."
ORIGIN

High quality sequence stop: 697.

Query Match   18.1%;   Score 573;   DB 2;   Length 852;
Best Local Similarity   91.3%;   Prd. No. 7.4e-137;
Matches 742;   Conservative   0;   Mismatches 55;   Indels 16;   Gaps   12;

```
Qy 1495 TTGGATAAGTAAAATGAGATAATTTTC-AAATCATCATAGAATTGATTTGAGCCAGCAGGG  1553
        |||||||||||||||||||||||||||  |||||||||||||||||||||||||||||||
Db    1 TTGGATAAGTAAAATGAGATAA-TTTTCAAAATCATCATAGAATTGTTTTGAGCCAGCAGGG    60

Qy 1554 CTATCCATGATGCTTCCTTGTTCTATCCATGATGAGCAGCCATTCCACCTTTATCAGAGTGG  1613
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CTATCCATGATGGCTTCCTTGTTCTATCCATGATGAGCAGCCATTCCACCTTTATCAGAGTGG  120

Qy 1614 CAAAGACCCTTC-CAATGTGTAATGAAACTGTG-CAATGTGTCTGAATCATGCG  1672
        ||||||||||||  ||||||||||||||||||| |||||||||||||||||||||
Db  121 CAAAGACCCTTTCCAATGTGTAATGAAACTGTGTCAATGTGTCTGAATCATGCG  180

Qy 1673 CCGAGTTTTTACACAGCTCTTGATGATTCAGGAGAAACAAAATACCCTCTTTTTTTTACTTAGCAA  1732
        |||||   |||||||||||||| |||||| |||||||||||||||| ||  |||||||||||||
Db  181 CCGAG--TTTTACACAGCATTTTCTCCGAG-TCAGGAGAAAAGTC---TTTATTAGCAA  239

Qy 1733 AGAGTACAGAAAACAAAATACCCTGTGCCTGCTTCTGCTTCAGTTTGTGACATTAAGCTGGGTTTGGTTGGA  1792
        |||||||||||||||||||||||| |||||||||||||||||||| ||||||||||||||||||||
Db  240 AGAGTACAGACAAACAAAAATACCCTGTGCCTGCTTCTGCTTCAGTTTGTGACATTAAGCTGGGT-TTGGTTGGA  295

Qy 1793 TGTGCCTGCTTCTGCTTCAGTTTTTGGGGCGTTTCTTCCTTACCTTTATCAGTTTGTATTC  1852
        ||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db  296 TGTGCCTGCTTCTGCTTCAGTTTTTGGGGCGTTTCTTCCTTATCAGTTTGTATTC  354

Qy 1853 TTGTCTATATTTCCTTCACTTCTCAATTCTCTCTCCCACTGACAGGATCAACTCAATGAC  1912
        |||||||||||||||||||||||  |||||||||||||||||||||||||||||||||||
Db  355 TTGTCTATATTTCCTTCACTTCTCAATTCTCTCTCCCACTGACAGGATCAACTCAATGAC  414

Qy 1913 GAGCTTCCTGCTTTGGGATTCTGCTTCTCACATGTTTTATCCATAAAGTTACTCATCT  1972
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  415 GAGCTTCCTGCTTTGGGATTCTGCTTCTCACATGTTTTATCCATAAAGTTACTCATCT  474

Qy 1973 ATAAAGTAGTTC-AAACATCCATGCTTCTCACATGTATTATCCATAAAGTTACTCATCT  2031
        |||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||
Db  475 ATAAAGTAGTTCAAAAACATCCATGCTTCTCACATGTATTATCCATAAAGTTACTCATCT  534

Qy 2032 GATTTTATTTAAATAGTGAACTCTACTTGA-TATCAGACCCGAGGACCATCCTCCATT  2090
        |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db  535 GATTTTATTTAAATAGTGAACTCTACTTGATTATCAGACCCGAGGACCATCCTCCATT  594

Qy 2091 GGAATAATGAAGATATTGTCATGGCAGAAAAGCAGGTGTGCCATTA--ATTGATAA  2148
        |||||||||||||||||||| |||||||||||||||||||||||||   |||||||
Db  595 GGAGAATATGAAGATATTGTCATGGCAGAAAAGCAGGTGTGCCATTAATATTGATAA  654

Qy 2149 GATACC-ACAAGCATCATCATGCAGTTATGAACACAGTGCTGAAA-GGATCATAGAC--  2204
        |||||| |||||||||||||||  |||||||||||||||||||||| ||||||||||
Db  655 GATACCAACAAGCATCATCATGCAGTATGAACACCAGTGCTGAAACGGATCATAAACCA  714

Qy 2205 AGGGGTGGTTAAATCTGATCCCGTAGAATAAAACTTCAGTGTACCTATTTCAGGGAAGAG  2264
```

```
Db    239  CTCCCCCAATCACAGGGAGGCG TAGTAAGATG-TGGAAGCAAACAAATCTTGTGACT  297
                                     |||||||||||  |||||||||||||||||||||||
Db    715  GGGGTCGGTTAAATTCTGATCCAGTAGAATAACTTCAGGGTACCTATTTCAAGGAAAGAA  774

Qy    779  GTCAATCCCAGCATTATAACCC GAAGTACCAGAATTGAAGGAGGAAGAAAGCACAGTTT  838
                                   |||||||||  ||||||||||||||||||||||||
Qy    2265 TTAATTTCACAATTAAAACTAGTAAATGAACCA  2297

Db    298  GTCAATCCCAGCATTATAACCC GAAGTACCAGAA-TTGAAGGAGGAAGAAAGCACAGTTT  356
                                   |||||||||  ||||||||  ||||||||||||||
Db    775  TTACAGTCCCATTAACAACAATAGTAAATGACCCA  807

Qy    839  TTGGAGGACATCTCCACCTATG AGATGCATTCTCCTCCGCCAGCATTTTCCTATCGG  898

RESULT 15
BI100002
Qy    357  TTGGAGGACATCTCCACCTATG AGATGCATTCTCCTCCGCCAGCATTTTCCTATCGG  416

LOCUS         BI100002                 771 bp     mRNA     linear   EST 26-JUN-2001
Db    899  GCCAACACCAGGCATCTCTTTA AGTTACCAAACACTCAAAGAGTCAAAAATGAGGCAA  958
DEFINITION    602884742F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:5039963
                                   |||||||||  ||||||||||||||||||||||||
              5', mRNA sequence.
Db    417  GCCACACCAGGCATCTCTTTA AGTCTACCAAACACTCAAAGAGTCAAAAATGAGGCAA  476
ACCESSION     BI100002
VERSION       BI100002.1  GI:14550895
Qy    959  AAGGTTCTTCTTCCTCCATCCCA GTACCTGAGACACCTTCGCTCTTTTCTGTGGAACT-AA  1017
KEYWORDS      EST.
SOURCE        Mus musculus (house mouse)
Db    477  AAGGTTCTTCTTCTTCCTCCATCCCA GTACCTGAGACACCTTCGCTCTTTTCTGTGGAACT  536
  ORGANISM    Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Qy    1018 AGGGGTGACTGCATACCGCTTG CCACAGGCTTGATGATTGCAAGTGTCGTGTGGAACT  1077
              Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
              Sciurognathi; Muroidea; Muridae; Murinae; Mus.
Db    537  AGGGGTGACTGCATACCGCTTG CCACAGGCTTGATGATTGCAAGTGTCGTGTGGAACT  596
REFERENCE     1  (bases 1 to 771)
  AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
Qy    1078 GTGTGAAAACGTGAAGCTTCTAC GATTCTCGGCCTTCTCTAAGACTATCGAAGACACCCC  1137
  TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL     Unpublished (1999)
Db    597  GTGTG-AAACGTGAAGCTCTA- GATTCTCGGCCCTTCTCTAAGA-TATCGAAGAGCGC-  652
COMMENT       Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
Qy    1138 ACTCAGTCACCACTACTATGAT ACATGTTACCTAAGCATGGTTTCCACCAGATGCCTAA  1197
              Tissue Procurement: Jeffrey E. Green, M.D.
              cDNA Library Preparation: Life Technologies, Inc.
Db    653  -ATCGTCACCA-TATTATGAT CAGTGTTACTAAGCATGGTTTCCACGAGATGCCT-A  709
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
Qy    1198 AGAATACAGCCAAATGCTCCAG TCCATATGAGAGGAATC  1237
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
Db    710  AGAATACAGGCAAATGATGGAG TCCATATGAGAGGAATC  748
              http://image.llnl.gov
              Plate: LLAM11110 row: d column: 12
FEATURES      High quality sequence stop: 645.
Search completed: May 31, 2006, 19:4 :37          Location/Qualifiers
Job time : 14325 secs                     source          1..771
                                                          /organism="Mus musculus"
                                                          /mol_type="mRNA"
                                                          /strain="FVB/N"
                                                          /db_xref="taxon:10090"
                                                          /clone="IMAGE:5039963"
                                                          /lab_host="DH10B (T1 phage-resistant)"
                                                          /clone_lib="NCI_CGAP_Kid14"
                                                          /note="Organ: kidney Vector: pCMV-SPORT6; Site_1: NotI;
                                                          Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                                                          Average insert size 1.75 kb. Constructed by Life
                                                          Technologies. Note: this is a NCI_CGAP Library. |"
ORIGIN

Query Match     18.0%;    Score 570;  DB 2;  Length 771;
Best Local Similarity   95.5%;   Pred. No. 4.3e-136;
Matches  726;  Conservative   0;  Mismatches  20;  Indels  14;  Gaps  13;

Qy    480  CCATTCAAGACTTCGTGGTTTCCCAGAACAACACTCCAGTGGGACTAACATGAGCTACG  539
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1    CCATTCAAGACTTCGTGGTTTCCCAGAACAACACTCCAGTGGGACTAACATGAGCTACG  60

Qy    540  AGGTGGAAAGCAAGAAAACAACATCCCATTCGAGAGAACATTTTCCACA-TGTTTCCAGTG  598
           ||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db    61   AGGTGGAAAGCAAGAAAACAACATCCCATTCGAGAGAACATTTTCCACAGTGTTTCCAGTG  120

Qy    599  TCGCAGCCTTTTGTGGACTATCCTATAACCAGTGTGCAGTGGTTAATGGGGGAATT  658
           |||||||||  ||||||||||||||||||||||||||||||||||||||||||||||
Db    121  TCGCAGCC-TTTGTGGACTATCCTATAACCAGTGTGCAGTGGTTAATGGGGGAATT  179

Qy    659  CTCAACAAGTCTCTCGCGGAGCAGAGAAATTGATAAATCGACTTCGTTCTTCAGGTGTAAC  718
           ||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db    180  CTCAACAAGTCTCTCGCGGAGCAGAGAAA-TGATAAATCGACTTCGTTCTTCAGGTGTAAC  238

Qy    719  CTCCCCCAATCACAGGGAGCGCGCTAGTAAAGATGTTGGAAGCAAAACAAATCTTGTGACT  778
           ||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
```

OM nucleic - nucleic search, using sw model

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

Run on: May 31, 2006, 10:17:18 ; Search time 1881 Seconds
                                          (without alignments)
                                          11735.321 Million cell updates/sec

Title: US-10-501-930-2
Perfect score: 3166
Sequence: 1 cggagcggcgagtcggtgcc.........gctattagcaaaaaaaa 3166

Scoring table: IDENTITY_NUC
                  Gapop 10.0 , Gapext 1.0

Searched: 524920 seqs, 348612231 residues

Total number of hits satisfying chosen parameters: 10489840

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Database :   N_Geneseq_8:*
                  1: geneseqn1980s:*
                  2: geneseqn1990s:*
                  3: geneseqn2000s:*
                  4: geneseqn2001as:*
                  5: geneseqn2001bs:*
                  6: geneseqn2002as:*
                  7: geneseqn2002bs:*
                  8: geneseqn2003as:*
                  9: geneseqn2003bs:*
                 10: geneseqn2003cs:*
                 11: geneseqn2003ds:*
                 12: geneseqn2004as:*
                 13: geneseqn2004bs:*
                 14: geneseqn2005s:*
                 15: geneseqn2006s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID		Description
1	3166	100.0	3166	10	ADC73012	Adc73012 Murine ST
2	930.2	29.4	1500	10	ADC73014	Adc73014 Human ST8
3	744.8	23.5	2008	4	AAH77592	Aah77592 Human sia
4	614.6	19.4	840	13	ADV69200	Adv69200 Human cel
5	224.4	7.1	1071	6	ABV78252	Abv78252 Human GD3
6	224.4	7.1	1071	6	ABZ35828	Abz35828 Human GD3
7	224.4	7.1	1071	6	ABX10071	Abx10071 Human GD3
8	224.4	7.1	1071	6	ABL91793	Abl91793 Human pol
9	224.4	7.1	2117	2	AAQ77831	Aaq77831 Human alp
10	224.4	7.1	2117	13	ADR24998	Adr24998 Breast ca
11	222.2	7.0	1704	2	AAT03694	Aat03694 Human alp
12	207	6.5	1223	10	ADB58122	Adb58122 Toxicity-
13	175.4	5.5	1511	4	AAH34115	Aah34115 Human col
14	131.4	4.2	448	4	ABA56318	Aba56318 Human foe
15	131.4	4.2	448	4	ABA45804	Aba45804 Human bre
16	131.4	4.2	448	4	AAK29997	Aak29997 Human bon
17	131.4	4.2	448	6	ABS04570	Abs04570 Human gen
18	130.8	4.1	428	4	ABA68901	Aba68901 Human foe

19	130.8	4.1	428	4	ABA50923		Aba50923 Human bre
20	130.8	4.1	428	4	ABA35853		Aba35853 Probe #14
21	130.8	4.1	428	4	AAK43020		Aak43020 Human bon
22	130.8	4.1	428	4	AAK17233		Aak17233 Human bra
23	130.8	3.5	428	6	ABS17087		Abs17087 Human gen
24	112.2	2.6	150	6	ABQ99173		Abq99173 Human ORF
25	81.8	2.6	1426	10	ADB59119		Adb59119 Toxicity-
26	79.6	2.5	1048	2	AAT39664		Aat39664 Extracell
27	79.6	2.5	1660	2	AAT39663		Aat39663 Sia alpha
28	79.6	2.5	1660	6	ABI99514		Abi99514 Mouse isc
29 c	79.4	2.5	293	6	ABQ99108		Abq99108 Human ORF
30	75	2.4	472	9	ACH42114		Ach42114 Human foe
31	58	1.8	2000	8	ADA71938		Ada71938 Rice gene
32	53.6	1.7	2131	14	ADZ71009		Adz71009 Human chr
33	53.2	1.7	494	9	ACH13893		Ach13893 Human adu
34 c	52.6	1.7	542	12	ACH80371		Ach80371 Human gen
35	52.6	1.7	601	6	ABQ47664		Abq47664 Oligonucl
36 c	52.6	1.7	601	6	ABQ47665		Abq47665 Oligonucl
37	52.2	1.6	15416	6	ABL14231		Abl14231 Human imm
38	52.2	1.6	15416	6	ABK31527		Abk31527 Signal tr
39	52.2	1.6	15416	6	ABL70498		Abl70498 Chemicall
40	52.2	1.6	15416	6	AAS61453		Aas61453 Human gen
41	51.6	1.6	1993	14	ADZ15753		Adz15753 DNA encod
42	51.2	1.6	1894	14	AEC82858		Aec82858 Breast ca
43	51.2	1.6	10865	13	ADS89721		Ads89721 Oligonucl
44 c	50.6	1.6	581	12	ACH68308		Ach68308 Human gen
45	50.6	1.6	1288	9	ACF05972		Acf05972 Human car

ALIGNMENTS

RESULT 1
ADC73012
ID   ADC73012 standard; cDNA; 3166 BP.
XX
AC   ADC73012;
XX
DT   01-JAN-2004 (first entry)
XX
DE   Murine ST8Sia VI cDNA.
XX
KW   O-glycan alpha2,8-sialyltransferase;
KW   beta-galactoside alpha2,6-sialyltransferase; cytostatic; virucide;
KW   antiinflammatory; neuroprotective; cancer metastasis; viral infection;
KW   inflammation; nerve tissue; murine; mouse; ST8Sia VI; gene; ss.
XX
OS   Mus sp.
XX
FH   Key   Location/Qualifiers
FT   CDS   77..1273
FT   /*tag= a
FT   /product= "Murine ST8Sia VI protein"
XX
PN   WO2003064655-A1.
XX
PD   07-AUG-2003.
XX
PF   30-JAN-2003; 2003WO-JP000883.
XX
PR   30-JAN-2002; 2002JP-00021159.
PR   24-APR-2002; 2002JP-00122673.
XX
PA   (RIKE) RIKEN KK.
XX
PI   Takashima S, Tsujimoto M, Tsuji S;
XX
DR   WPI; 2003-627613/59.
DR   P-PSDB; ADC73011.
XX
PT   Sugar-chain synthases which are sialyltransferases and encoded genes,
PT   applicable in drugs for inhibiting cancer metastasis, preventing viral
PT   infection, inhibiting inflammation and potentiating nerve tissues.

XX
PS   Claim 2; SEQ ID NO 2; 97pp; Japanese.
XX
CC   The invention relates to a novel O-glycan alpha2,8-sialyltransferase
CC   having a novel substrate specificity and selectivity and a novel beta-
CC   galactoside alpha2,6-sialyltransferase having a novel substrate
CC   specificity and selectivity. The enzymes of the invention demonstrate
CC   cytostatic, virucide, antiinflammatory and neuroprotective activities and
CC   may be applicable in drugs for inhibiting cancer metastasis, preventing
CC   viral infection, inhibiting inflammation and potentiating nerve tissues.
CC   The current sequence is that of the murine ST8Sia VI cDNA of the
XX   invention.
SQ   Sequence 3166 BP; 937 A; 706 C; 645 G; 878 T; 0 U; 0 Other;

Query Match   100.0%; Score 3166; DB 10; Length 3166;
Best Local Similarity   100.0%; Pred. No. 0;
Matches 3166; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy      1 CGGAGCGGCGAGTCGGTGCCGCCCGGGCTGCGCTTCGCCCCGGCAGCTTGGCGGCGGAGG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CGGAGCGGCGAGTCGGTGCCGCCCGGGCTGCGCTTCGCCCCGGCAGCTTGGCGGCGGAGG 60

Qy     61 ACGCCCGTGGCTCAGGATGAGATCGGGGGGCACGCGTGTTCGCCCTCATAGGCAGCCTGAT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ACGCCCGTGGCTCAGGATGAGATCGGGGGGCACGCGTGTTCGCCCTCATAGGCAGCCTGAT 120

Qy    121 GCTGCTGCTCCTCCTCCTGCGTATGCTCCTGTGTGCCCAGCCGCCTCGTGCCCCGCTTCGCCCAGGCT 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GCTGCTGCTCCTCCTCCTGCGTATGCTCCTGTGTGCCCAGCCGCCTCGTGCCCCGCTTCGCCCAGGCT 180

Qy    181 GTTGATGGAGGGAAGCAGAGGAGGACACCAGTGTACCTCAGCTGCACTGAAGACACTCTG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GTTGATGGAGGGAAGCAGAGGAGGACACCAGTGTACCTCAGCTGCACTGAAGACACTCTG 240

Qy    241 GAGCCCGACAACCCCGGTACCACGCACCAGGAACAGCACATATCTGATGAGAAGACAAAC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GAGCCCGACAACCCCGGTACCACGCACCAGGAACAGCACATATCTGATGAGAAGACAAAC 300

Qy    301 CCAAATAACAGAGAAATGCAAAGATCTGCAATATAGCTTGAACTCTTTATCTCAACAAAAC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CCAAATAACAGAGAAATGCAAAGATCTGCAATATAGCTTGAACTCTTTATCTCAACAAAAC 360

Qy    361 GAGACGGTACTCTGAGGATGACTACCTCCAGACCATCACAAACATCAGAGATGCCCATG 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GAGACGGTACTCTGAGGATGACTACCTCCAGACCATCACAAACATCAGAGATGCCCATG 420

Qy    421 GAACCGGCAAGCAAGACGAAGAATGAGCCAGTGGGGCTAACATGCTAACATGAGCTACGA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GAACCGGCAAGCAAGACGAAGAATGAGCCAGTGGGGCTAACATGCTAACATGAGCTACGA 480

Qy    481 CATTCAAGACTTCGTGGTTCCCAGAACAACACTCGAGAGAACATTTTCCACATGTTCCAGTGTC 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CATTCAAGACTTCGTGGTTCCCAGAACAACACTCGAGAGAACATTTTCCACATGTTCCAGTGTC 540

Qy    541 GGTGGAAAGCAAGAACACATCCCATTCGAGAGAACATTTTCCACATGTTCCAGTGTC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GGTGGAAAGCAAGAACACATCCCATTCGAGAGAACATTTTCCACATGTTCCAGTGTC 600

Qy    601 GCAGCCTTTTGTGGACTATCCCTATAACCAGTGTGCAGTTGGTAATGGGGGAATTCT 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GCAGCCTTTTGTGGACTATCCCTATAACCAGTGTGCAGTTGGTAATGGGGGAATTCT 660

Qy    661 CAACAAGTCTCTCTGCCGGAGACGGCGCTAGTAAAGATGTTGGAAGCAAAACAAATCTTGTGACTGT 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 CAACAAGTCTCTCTGCCGGAGACGGCGCTAGTAAAGATGTTGGAAGCAAAACAAATCTTGTGACTGT 720

Qy    721 CCCCCAATCACAGGAGACGGCTAGTAAGAGATGTTGGAAGCAAAACAAATCTTGTGACTGT 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 CCCCCAATCACAGGAGACGGCTAGTAAGAGATGTTGGAAGCAAAACAAATCTTGTGACTGT 780

Qy    781 CAATCCCAGCATTATAACCCTGAAGTACCAGAATTGAAGGAGAGAAAGCACAGTTTTT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 CAATCCCAGCATTATAACCCTGAAGTACCAGAATTGAAGGAGAGAAAGCACAGTTTTT 840

Qy    841 GGGAGGACATCTCCACCTATGGAGATGCATTCCTCCTGCCAGCATTTCCTATCGGGC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 GGGAGGACATCTCCACCTATGGAGATGCATTCCTCCTGCCAGCATTTCCTATCGGGC 900

Qy    901 CAACACAGGCATCTCTTTAAACTACCAAAACACTCAAAGAGTCAAAAATGAGGCAAA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 CAACACAGGCATCTCTTTAAACTACCAAAACACTCAAAGAGTCAAAAATGAGGCAAA 960

Qy    961 GGTTCTTCTTCTTCCATCCCAGGACCTGAGACACACCTCGCTCTTTTCTGGAGAACTAAAGG 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 GGTTCTTCTTCTTCCATCCCAGGACCTGAGACACACCTCGCTCTTTTCTGGAGAACTAAAGG 1020

Qy   1021 GGTGACTGCATACCGCTGTCCCAGGGCTTGATTGCAAGTGTCGCTGTGGAACTGTG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 GGTGACTGCATACCGCTGTCCCAGGGCTTGATTGCAAGTGTCGCTGTGGAACTGTG 1080

Qy   1081 TGAAAACGTGAAGCTCTACGGATCTGGCCTTTCTCTAAGACTATCGAAGACACCCCACT 1140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 TGAAAACGTGAAGCTCTACGGATCTGGCCTTTCTCTAAGACTATCGAAGACACCCCACT 1140

Qy   1141 CAGTCACCACTACTATGATAACGTTACCTAAGCATGGTTCCACCAGATGCCTAAAGA 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CAGTCACCACTACTATGATAACGTTACCTAAGCATGGTTCCACCAGATGCCTAAAGA 1200

Qy   1201 ATACAGCCAAATGCTCCAGTCCGATATGAGGAGGAATAATTCAGGAGGTGGAGATGGTGTCA 1260
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 ATACAGCCAAATGCTCCAGTCCGATATGAGGAGGAATAATTCAGGAGGTGGAGATGGTGTCA 1260

Qy   1261 TGAAACGGCTTAACGTTCTTAGAGAAACGGCTTAACGTTCTTAGAGAAACGGCTTAACGTTCTTA 1320
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 TGAAACGGCTTAACGTTCTTAGAGAAACGGCTTAACGTTCTTAGAGAAACGGCTTAACGTTCTTA 1320

Qy   1321 CAGCATCCCAAAAAGCCAATACAAGAGGCACAGAGAAAGCATGAATTACAAGGCGCT 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 CAGCATCCCAAAAAGCCAATACAAGAGGCACAGAGAAAGCATGAATTACAAGGCGCT 1380

Qy   1381 CTCCCACTTGTCTAGACCAAAGCACCCGCCCCCAACTCACTTTGCAGCCTCCACGAGTCA 1440
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 CTCCCACTTGTCTAGACCAAAGCACCCGCCCCCAACTCACTTTGCAGCCTCCACGAGTCA 1440

Qy   1441 CTCATTCTCACCTTCAAGTTCCAAAACATCAGACTTGGAT 1500
          ||||||||||||||||||||||||||||||||||||||||
Db   1441 CTCATTCTCACCTTCAAGTTCCAAAACATCAGACTTGGAT 1500

Qy   1501 AAGTAAAATGAGATAATTTTCAATCATCAGAGTGGCCAGGGGTCTCTCA 1560
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 AAGTAAAATGAGATAATTTTCAATCATCAGAGTGGCCAGGGGTCTCTCA 1560

Qy   1561 GAATGCTTCCTTGTCCTATCCGATAGCCATTCCCACCTTTATCAGAGTGGTAATGAA 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 GAATGCTTCCTTGTCCTATCCGATAGCCATTCCCACCTTTATCAGAGTGGTAATGAA 1620

Qy   1621 ACTGTGCAATTGTGCCAAAGACCTTTCTGAAGGAATGTCTGAATCATGCGCCCAGTTT 1680
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 ACTGTGCAATTGTGCCAAAGACCTTTCTGAAGGAATGTCTGAATCATGCGCCCAGTTT 1680

Qy   1681 TTACACACAGCTCTCCTTTATAATAAATCCTTCCCATTCTCCCCTCCCTCCTAGTAGTACA 1740
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 TTACACACAGCTCTCCTTTATAATAAATCCTTCCCATTCTCCCCTCCCTCCTAGTAGTACA 1740

Qy   1741 GAAACAAAATACCCTTGATGATCAGGAAGAAAGTCTTTTTTACTTAGCAATGTGCCTG 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 GAAACAAAATACCCTTGATGATCAGGAAGAAAGTCTTTTTTACTTAGCAATGTGCCTG 1800

Qy   1801 CTTCTGATTCAGTTCGCTTCGTGCATTAAGCTGGGTTTGGGTTTTGGGGGC 1860
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 CTTCTGATTCAGTTCGCTTCGTGCATTAAGCTGGGTTTGGGTTTTGGGGGC 1860

Qy   1861 GTTTCTTCACTTCTTTTGTCTATTTCCTTATCAGTTTGATTCGAGCTTCC 1920
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 GTTTCTTCACTTCTTTTGTCTATTTCCTTATCAGTTTGATTCGAGCTTCC 1920
```

```
Qy   3001 ATTTTGCAAAGGGTAAAGGCTTTTTGTTAAATAAAATAAAATTATTATTTCTTCTG 3060
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 ATTTTGCAAAGGGTAAAGGCTTTTTGTTAAATAAAATAAAATTATTATTTCTTCTG 3060

Qy   3061 ATGAATAGAGGCTCCTTTATGCGCTGCTAATGAACCTAATTAGCTTAAATATCTCCT 3120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 ATGAATAGAGGCTCCTTTATGCGCTGCTAATGAACCTAATTAGCTTAAATATCTCCT 3120

Qy   3121 AGCAACATTGGTCACGTTTCAACATGCTATTAGCAAAAAAAAAAA 3166
          |||||||||||||||||||||||||||||||||||||||||||||
Db   3121 AGCAACATTGGTCACGTTTCAACATGCTATTAGCAAAAAAAAAAA 3166
```

RESULT 2
ADC73014

```
ID   ADC73014 standard; cDNA; 1500 B.
XX
AC   ADC73014;
XX
DT   01-JAN-2004 (first entry)
XX
DE   Human ST8Sia VI cDNA.
XX
KW   O-glycan alpha2,8-sialyltransferase;
KW   beta-galactoside alpha2,6-sialyltransferase; cytostatic; virucide;
KW   antiinflammatory; neuroprotective; cancer metastasis; viral infection;
KW   inflammation; nerve tissue; human; ST8Sia VI; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             92..1288
FT                   /tag= a
FT                   /product= "Human ST8Sia VI protein"
XX
PN   WO2003064655-A1.
XX
PD   07-AUG-2003.
XX
PP   30-JAN-2003; 2003WO-JP000883.
XX
PR   30-JAN-2002; 2002JP-00021159.
PR   24-APR-2002; 2002JP-00122673.
XX
PA   (RIKE ) RIKEN KK.
XX
PI   Takashima S, Tsujimoto M, Tsuda S;
XX
DR   WPI; 2003-627613/59.
DR   P-PSDB; ADC73013.
XX
PT   Sugar-chain synthases which are sialyltransferases and encoded genes,
PT   applicable in drugs for inhibiting cancer metastasis, preventing viral
PT   infection, inhibiting inflammation and potentiating nerve tissues.
XX
PS   Claim 4; SEQ ID NO 4; 97pp; Japanese.
XX
CC   The invention relates to a novel O-glycan alpha2,8-sialyltransferase
CC   having a novel substrate specificity and selectivity and a novel beta-
CC   galactoside alpha2,6-sialyltransferase having a novel substrate
CC   specificity and selectivity. The enzymes of the invention demonstrate
CC   cytostatic, virucide, antiinflammatory and neuroprotective activities and
CC   may be applicable in drugs for inhibiting cancer metastasis, preventing
CC   viral infection, inhibiting inflammation and potentiating nerve tissues.
CC   The current sequence is that of the human ST8Sia VI cDNA of the
CC   invention.
XX
SQ   Sequence 1500 BP; 425 A; 357 C; 349 G; 369 T; 0 U; 0 Other;

Query Match             29.4%;  Score 930.2; DB 10;  Length 1500;
Best Local Similarity   78.4%;  Prob No. 2.1e-233;
Matches 1169;  Conservative 0;  Mismatches 308;  Indels 14;  Gaps 4;
```

```
Qy   1921 TGCTTTGGGATTCTGCAATTCTCTCTCCCACTGACAGGATCAACTCAATGACATAAAGTA 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 TGCTTTGGGATTCTGCAATTCTCTCTCCCACTGACAGGATCAACTCAATGACATAAAGTA 1980

Qy   1981 GTTCAAACATCCATTGCTTCTCCTCCACATGTTTATCCATAAGTTACTCATCTGATTTATT 2040
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 GTTCAAACATCCATTGCTTCTCCTCCACATGTTTATCCATAAGTTACTCATCTGATTTTATT 2040

Qy   2041 TAAAATAGTGAACATCTACTTGATATCAGACCCGAGGACCATCCTCCATTGGAGAATATG 2100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 TAAAATAGTGAACATCTACTTGATATCAGACCCGAGGACCATCCTCCATTGGAGAATATG 2100

Qy   2101 AAGATATTGTCACTGGCAGAAAAGCAGGTGTGTGCCATTAATTGATAAGATACCACAAGC 2160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 AAGATATTGTCACTGGCAGAAAAGCAGGTGTGTGCCATTAATTGATAAGATACCACAAGC 2160

Qy   2161 ATCATCATGCCAGTTATGAACACAGTGCTGAAAGGATCATAGACAGGGTGTTAAATCT 2220
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 ATCATCATGCCAGTTATGAACACAGTGCTGAAAGGATCATAGACAGGGTGTTAAATCT 2220

Qy   2221 GATCCCAGTAGAATAAACTTCAGTGTACCTATTCAGGGAAGAGTTAATTTCACAATTAA 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 GATCCCAGTAGAATAAACTTCAGTGTACCTATTCAGGGAAGAGTTAATTTCACAATTAA 2280

Qy   2281 AACTAGTAAATGAACCAATTCTTAGGCACATTAAGTGGATTCGAGTAAAAGAAAGGGAA 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 AACTAGTAAATGAACCAATTCTTAGGCACATTAAGTGGATTCGAGTAAAAGAAAGGGAA 2340

Qy   2341 CAGCAGGAGAAAAGCTGTTCGCTTGGTTCTGATTACCCAAATGAGCATGCTGGAAGGAGGT 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 CAGCAGGAGAAAAGCTGTTCGCTTGGTTCTGATTACCCAAATGAGCATGCTGGAAGGAGGT 2400

Qy   2401 TGTGAGGCTACGCTAAAACCTCGCGTAGGGAGAGAGTACAGTGCATGAGTGTGGCGCT 2460
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 TGTGAGGCTACGCTAAAACCTCGCGTAGGGAGAGAGTACAGTGCATGAGTGTGGCGCT 2460

Qy   2461 TTTGTCCACACTCGTGAAGGGTGAGTAATTCAGAGCCAATCACATCACAAGGATGGACAC 2520
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 TTTGTCCACACTCGTGAAGGGTGAGTAATTCAGAGCCAATCACATCACAAGGATGGACAC 2520

Qy   2521 ACCTAACTCATCACTTCAGGGGGAGATGAATGCTTCATGAGAAATTACACTCATAAGCT 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 ACCTAACTCATCACTTCAGGGGGAGATGAATGCTTCATGAGAAATTACACTCATAAGCT 2580

Qy   2581 AAGCACAGTTTGAGTAAAATTGAGTGATGTTAAATATGAACATTTTATACCTCTTA 2640
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 AAGCACAGTTTGAGTAAAATTGAGTGATGTTAAATATGAACATTTTATACCTCTTA 2640

Qy   2641 CTAATGTCCCACCGACACCTTTTAAATGTAAGCACATTTATTATTAAGTTACTTGACATT 2700
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 CTAATGTCCCACCGACACCTTTTAAATGTAAGCACATTTATTATTAAGTTACTTGACATT 2700

Qy   2701 AAATGCTTATGTCTGTATATTCTGTTCATCCATCGATTTTCCCAAAAAGTAAGAGCATAG 2760
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 AAATGCTTATGTCTGTATATTCTGTTCATCCATCGATTTTCCCAAAAAGTAAGAGCATAG 2760

Qy   2761 GAGATGAGGCCTACATGCCAAGAAAACTATAAATTTTACTCTTTAATTCTTGAGCC 2820
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 GAGATGAGGCCTACATGCCAAGAAAACTATAAATTTTACTCTTTAATTCTTGAGCC 2820

Qy   2821 AGCTTGTTGTTTATCAAGTGCTTTTTGAAGAGACAGCACCCTGAATTCTTCATTCTG 2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 AGCTTGTTGTTTATCAAGTGCTTTTTGAAGAGACAGCACCCTGAATTCTTCATTCTG 2880

Qy   2881 ATACAGTGTCACCTTGTATTTAACATTTGTAATGTGTTTCAAGTTTACATCTCTTTCAT 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 ATACAGTGTCACCTTGTATTTAACATTTGTAATGTGTTTCAAGTTTACATCTCTTTCAT 2940

Qy   2941 TCTTTTATAGCAAATCAAACGTATTAGCTTCAGAAATTTATCAGAAGTTCATATATAAAT 3000
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 TCTTTTATAGCAAATCAAACGTATTAGCTTCAGAAATTTATCAGAAGTTCATATATAAAT 3000
```

```
Qy      2  GGAGCGGCGAGTCGGTGCCGCCCGGGCGCTGCGCTTCGCCGGCGAGCTTGGCGGCGAGGA   61
                                                                          
Db     17  GGAGCCCGCGAGTCGGGCGGCCGGGGCGGCCCCGGGGCGTGCTTCGCCGCGCCCGGCGGG   76

Qy     62  CGCCCGTGGCTCAGGATGAGATCGGGGGCGGCCGCTGTTCGCCCTGATCATAGGCAGCCTGATG  121
                                                                          
Db     77  CGCCTGTGGCTCAGGATGCGGCGGGGGCGCACTACGCTGCTCGCCGCCGCCAGCCTGCTG  136

Qy    122  CTGCTGCTCCTCCTGCGCTATGCTCTGGTGCCCAGCCAGCGCGCCTGCGCGCCTCCAGGCTG  181
                                                                          
Db    137  CTGCTGCTGCTGCTGCTGCTGCGTGCCCGCAGACGCGCCGGCGCCGGCGCCGCGCCAGGATT  196

Qy    182  TTGATGGGGGAAGCAGAGAGGACCAGTGGTACCTCAGCTGCACTGAAGACACTCTGG  241
                                                                          
Db    197  CTGGTGGAGGAAAGCCACCGGGTACCACGGCCACCCGGCGACCCCCAGCGCGCTGAGGACGCTCCGG  256

Qy    242  AGCCCGACAACCCCGGTACCGCGCCACCAGGAACACAGCACATATCTGGATGACGAAGACAACC  301
                                                                          
Db    257  AGCCCGGCGGTGGTACCGCGGGTACCGCGCCACTAACAGCACATATCTGAATGAGAAGTCGCTC  316

Qy    302  CAAATAACAGAGAGAAATGCAAAGATCTGCAAATTAGCTTGAACTCTTTATCTAACAAAACG  361
                                                                          
Db    317  CAACTGACGCGAGAAATGCAAAAATCTGCAATATGCATTGACTTGAGTCTTTCTCTAACAAAACG  376

Qy    362  AGACGGTACTCTGAGGATGACTACCTCCAGACCATCACAAACATACAGAGATGCCCATGG  421
                                                                          
Db    377  AAAGGGTATTCAGAGAACGACTACCTTCAGATTTATCACAGATTATACAGAGAGTTGTCCATGG  436

Qy    422  AACCGGCAAGCAAGCAGAGAAATATGACAATTTTAGAGCAAAACTGCTTCCTGTTGCGATGCC  481
                                                                          
Db    437  AAACGGCAAGCAAGCAGAGATATGCAAATTTAGAGCCAAACTTGCTTCCTGCTGTGATGCT  496

Qy    482  ATTCAAGACTTCGTGTTTCCAGAACAACACTCCAGTGGGGACTAACATGAGCTACGAG  541
                                                                          
Db    497  GTTCAAAACTTCGTGTTTCCAGAATAACACTCCAGTTGGGACTAAATGAGTTACGAG  556

Qy    542  GTGGAAAGCAAGAAAACACATCCCCATTCGAGAGAACATTTCCACATGTTTCCAGTGTCG  601
                                                                          
Db    557  GTGGAAAGCAAAAAAGAAATCCCAATTAAGAGAGAACATTTTCATATGTTTCCAGTGTCC  616

Qy    602  CAGCCTTTGTGGGACTATCCCTATAAACCCAGTGTGCAGTGGTTGGTAAAATCTGACTTCGTTCTC  661
                                                                          
Db    617  CAGCCTTTTGTGGGACTACCCTTATAAATCAGTGTGGTGCAGTGGTCGGAAATGGGGAATTCTG  676

Qy    662  AACAAGTCTCTCTGCCGGAGCAGAAATTGATGATAAATCTGACTTCGTGTCTTCAGGTGTAACCTC  721
                                                                          
Db    677  AATAAGTCTCTCTGTGGAGTAACTGAAATACTGACTTCGTTTTTAGGTGTAACCTA  736

Qy    722  CCCCCAATCACAGGAGCGCGTAGTAGACATGTTGGAAGCAAACAAATCTTGTGACTGTC  781
                                                                          
Db    737  CCCCCAACCACAGGAGAGTGTTAGTAGAGATGTTGGGCAGTAAAAACAAATCTTGTGACTGTC  796

Qy    782  AATCCCAGCATTATAACCCTGAAGTACCAGATTTGAAGGAGGAAGAAGCACAGTTTTTG  841
                                                                          
Db    797  AATCCAAGCATCATAACTCTGAAATATCTGGAACTAAAGAAAAAAAAGCCCTATTTTTG  856

Qy    842  GAGGACATCTCCACCTATGGAGATGCATTCCTCCCTGCCAGCATTTCCTGCCATCGGGCC  901
                                                                          
Db    857  GAGGACATTGCAACCTATGGAGATGCATTTTTTTTCTCGCCAGCATTTTCCTTCAGGGCC  916

Qy    902  AACACAGGCATCTCTTTTAAAGCTACCAAAACCTCAAAGAGTCAAAAATGAGGCAAAAG  961
                                                                          
Db    917  AACACGGTACCTCTTCTTCAAAGTATACTACACGCTCGAGAGTCTAAAGCAGCAGACAAAAG  976

Qy    962  GTTCTCTTCCATCCCAGTTACCTGAGACACCTGCTCTCTTTTCTGGAGAACTAAAGGG  1021
                                                                          
Db    977  GTTCTATTTTCCATCCCAAGTTACCTGAAAGATCTGGCCCTTTTCTGGAGAACTAAAGGT  1036

Qy   1022  GTGACTGCATCATACCGCTTGTCCACAGGCTTGATGATTGCGCTGTCGCTGTGAACTGT  1081
                                                                          
Db   1037  GTGACTGCGATACCGCTGTCGCCACGGCTTGATGATCACAAGTGTTGCAGTGGAACTGT  1096

Qy   1082  GAAAACTGAGAAGTCACGGATTCTCGGCCTTTCTCTAAGACTATCGAAGAACACCCCACTC  1141
                                                                          
```

```
Qy   1097  AAAAATGTGAAGCTGTATGGAT...CTGGCCCTTCTAAAACTGTAGAGACATACCTGTC  1156
                                                                          
Db  1142  AGTCACCACTACTATGATAACA...GTTACCTAAGCATGGTTCCACCAGTGCCTAAAGAA  1201

Qy  1157  AGCCATCACTATTATGACAACAG...AGTACCTAAACATGGTTTCCATCAGATGCCCAAAGAA  1216
                                                                          
Db  1202  TACAGCCAAATGTCCAGCTCCAGTCC...ATATGAGAGGAATCCTCAAACTGCAATTCAGCGAAATGT  1261

Qy  1217  TACAGCCAGATCCTCCAACTTCA...CATGAAAGGAATCCTCAAACTGCAATTAGCAAATGT  1276
                                                                          
Db  1262  GAAACGGCTT----AACGTTCT...TAGAAGAGGAGAATAATTTCAGGA-GGTGGAGTGGATGT  1316

Qy  1277  GAAGTCGCCTAAACAAAGTATC...TAAAATGGAATAATTTTAATATAATGCAGTAGGTGA  1336
                                                                          
Db  1317  GTCACAGCATCTCCAAAAAGCC...ATAGAAGAGAAGCACAGAGAAAGCATGAATTACAAAGG  1376

Qy  1337  TTAACAATGTCTCCAAACACC---...-----AAAGGAGGTGGCTAAAGAGTATTTTGAGATG  1388
                                                                          
Db  1377  CGCTCTCCCACTTGTCTAGACCA...AAGCCACCGCCCCCACTCACTTTGCAGCCTCCACGA  1436

Qy  1389  AGCCCCAAAATTTGGTTTGACCA...GAAGCTTCCCCACTCATTTTGAATGATGGCAAGTC-A  1447
                                                                          
Db  1437  GTCACTCATTCTCACCTTCAAC...GTCTTTCTTCTGAGAATAGAGACCAAAAC  1487

Qy  1448  TTCAATCCTTCTCATCTTCATT...TTCTCCCTTATAAACATGGACACCCATATC  1498
```

RESULT 3

AAH77592

ID   AAH77592 standard; cDNA; 2008 BP.
XX
AC   AAH77592;
XX
DT   22-OCT-2001 (first entry)
XX
DE   Human sialic transferase 27 cDNA.
XX
KW   Human; sialic transferase 27; cancer; cytostatic; immunological disease;
KW   human immunodeficiency virus; HIV; infection; inflammatory disease; ss.
XX
OS   Homo sapiens.
XX
PN   CN1298005-A.
XX
PD   06-JUN-2001.
XX
PF   29-NOV-1999; 99CN-00124142.
XX
PR   29-NOV-1999; 99CN-00124142.
XX
PA   (SHAN-) SHANGHAI BORONG GENE DEV CO LTD.
XX
PI   Mao Y, Xie Y, Qiu M;
XX
DR   WPI; 2001-489682/54.
DR   P-PSDB; AAG66954.
XX
PT   Human sialic transferase 27 as one new kind of polypeptide and
PT   polynucleotides encoding this polypeptide.
XX
PS   Claim 6; Page 21-22 (disclosure); 29pp; Chinese.
XX
CC   The invention relates to a novel polypeptide, human sialic transferase
CC   27, polynucleotides encoding this polypeptide and a DNA recombination
CC   process to produce the polypeptide. The polypeptide is useful for
CC   treating various diseases, such as malignant tumours, nosohaemia, HIV
CC   infection, immunological diseases and inflammatory diseases. The
CC   invention also provides an antibody directed against the polypeptide. The
CC   present sequence encodes the polypeptide of the invention
XX
SQ   Sequence 2008 BP; 631 A; 392 C; 351 G; 634 T; 0 U; 0 Other;

Query Match   23.5%; Score 744.8; DB 4; Length 2008;
Best Local Similarity   69.5%; Pred. No. 1.4e-184;
Matches 1292; Conservative 0; Mismatches 497; Indels 69; Gaps 18;

```
Qy    440  TATGACAATTTTAGAGCAAAACTGGCTTCCTGTTGCGATGCCATTCAAGACTTCGTCGTT  499
                ||||||||||||||||| |||||||||  |||| | |||||| | |||| ||| ||||
Db      1  TATGCAAATTTTAGAGCAAAACTGCTTCCTGCTGTGATGCTCTTCAAAACTTGTTGTT    60

Qy    500  TCCCAGAACAACACTCCAGTGGGGACTAACATGACTACGAGGTGGAAGCAAGAAACAC   559
           |||| |  |||||| |||||||| ||| | | || | || ||||  |||| |||||||||
Db     61  TCTCAGAATAACACTCCAGTGGGACTAATATGAGTTACGAGGTGGAAAGCAAAAAGAA   120

Qy    560  ATCCCCATTCGAGAGAGAACATTTTCCACATGTTTCCGAGCCTTTGTGGACTAT        619
           |||    ||| ||| |||| || |||||||  ||| ||||||||| | |||||||
Db    121  ATCCCAATTAAGAAGGAACAGCCATTTTCATTGTTTCCAGCCCTTTGTGGACTAC    180

Qy    620  CCCTATAACCAGTGTGCAGTGGTTGGTAATGGGGAATTCTCAACAAGTCTTCCTGCGGA   679
           || || |||||||| |||| ||| |||| ||||||||||||||||| |||||||| |||
Db    181  CCTTATAATCAGTGTGCAGTGGTTGGTAATGGGGAATTCTCAATAAGTCTCTCTGTGA   240

Qy    680  GCAGAAATTGATAAATCCGACTTCGTTTCAGTGTAACCTCCCCCAATCACAGGGAGC    739
           || |||||| ||| | ||| | ||| | ||||| ||| |||||||| |||||||||
Db    241  ACTGAAATAGATAAATCCGACTTCGTTTTAGTGTAACCTACCCCCAACCACAGAGAT   300

Qy    740  GCTAGTAAAGATGTTGGAAGCAAAACAAATCTGTGACTGTCAATCCCAGCATTATAACC   799
           || ||||||| ||||| |||||| | |||||| ||  |||||||| |||||||||| |
Db    301  GTTAGTAAAGATGTTGGCAGTAAAACAAATCTTGTGACTATAAAATCCAAGCATCATAACT   360

Qy    800  CTGAAGTACCCAGAATTGAAGGAGAAGAGCACAGTTTTGGAGGAGCATCTTCCACCTAT   859
           |||| ||||| || | |||| |||  |||||||||| |||||| |||||||  ||| |||
Db    361  CTGAAATATGGGAACTTAAAGGAAAAAAAAGCCCTATTTCTGGAGGAACATCAGGCAACCTAT   420

Qy    860  GGAGATGCATTCCTCCTGCCAGCATTTTCCTATGGGCCAACACAGGGCATCTCTTTT    919
           || ||||||||| | |  | || | | || | ||| | |||| |||  ||||  |||
Db    421  GGAGATGCATTTTTCTCTGCCTGCCCAGCAGGCATTTTAGTTGATTATCTGTGTCTCTTTC   480

Qy    920  AAAGTCTACCAAAACACTCAAAGTCAAAAATGAGGCAAAAGTTCTCCTTCCTCCATCCC   979
           ||| || ||| | | |   |||||  ||||  |||| ||| | ||  | |||||||||
Db    481  AAAGTATACACGGCTCGAAGAGTTCTGAGAACTAAAGCAAAAGTTTCTATTTTTCCATCCC   540

Qy    980  AGGTACCTGAGACACCCGCTCTTTTCTGAGAACTAAAGGGGTGACTCGCATACCGCTTG  1039
           || |||||||| |||| || |||| |||||| ||||| ||||| ||| |||||||||
Db    541  AAGTACCTGAAGATCTCGGCCCCTTTCTGGAGAACTAAAGGTGACTCGACTACCGCTTG   600

Qy   1040  TCCACAGGCTTGATGATTGCAAGTGCTGGCTGTGTGGGAACTGTGTGAAAACGTGAAGCTCTAC  1099
           || |||||||||  ||  ||| || || | |||| |||||||| | ||| ||| ||||||||
Db    601  TCCACCGGCTTGATGTCATCAAGTGTTGCAGTGGGAACTGTGTAAAAAATGTGAAGCTGTAT   660

Qy   1100  GGATTCTGGCCTTTCTCTAAGACTATGGAAGACACCCCCACTCAGTCACCACTACTATGAT  1159
           |||||| |||||||||  |||| |||||||| | || |||||| ||  || ||||  ||||
Db    661  GGATTCTGGCCCCTTTCTAAAACTGTAGAAGACATACCTGTCAGCCATCACTATTATGAC   720

Qy   1160  AACATGTTACCTAAGCATGGTTTCCACCAGATGCTCAAAGAATACAGCCAAATGTCCCAG  1219
           | |   ||||||||| ||||| | | | | |  ||||| ||||||||| || | |||||||
Db    721  AACAAGTTACCTAAACATGGTTTCCATCAGATATACAAAGAATACAGCCAAATACAGCCAGCCAA   780

Qy   1220  CTCCAATGAGAGGAATCCTCAAACTGCAATTCAGCAAATGTGAAACGGCTT----AACG  1275
           ||   | | |  ||||| |||||||| |||| ||||||| |||||| || |
Db    781  CTTCACATGACATGAAAGGAATCCTCAAACTGCAATTTAGCAAATGTGAAGTGCGCCTAAACAAAG   840

Qy   1276  TTTCTTAGAAGGAGAATAATTTCAGGA-GGTGGAGTGTGTCACAGCATCTCCAAAA  1334
           |   | | ||||| || || |||  || ||| ||  | |   || |||| | |||||
Db    841  TATCTTAAATGGGAATAATTTTAATAATCAGTAGGTGATTAATCATATGTCTCCAAAC   900

Qy   1335  AGCCAATAGAAGAGGCACAGAGAAAGCATGAATTACAAAGGCGCTCCCACTTGTCTA  1394
           |||||          |||| |  |||||| | || | || | ||||| || | |||
Db    901  ACC------AAAGGAGGTGGCTAAAGAGTATTTGAGATGAGCCCCAAAATTTGGTTT   952

Qy   1395  GACCAAAGCCACCCGCCCCCACTCACTTGCAGCCTCCACCAGTCACTC-ATTCTCACCT  1453
           ||||||     |||    |||| |  |||| | ||||| || |||||||||||| ||
Db    953  GACCAAAG------CTTCCCACTCATTTGCAATGATGGCAAGTCATTCATTCAATCCTTCTCA  1007

Qy   1454  TCAACGTTCTTTTCTCTGAGAATAGAGAACATCAGACTTGGATAAGTAAGTAAAATCAGA  1513
```

RESULT 4
ADV69200
ID   ADV69200 standard; DNA; 840 BP.
XX
AC   ADV69200;
XX
DT   24-FEB-2005 (first entry)
XX
DE   Human cell division regulating protein 27 encoding DNA SEQ ID NO 1.
XX
KW   cell division regulating protein 27; cancer; cytostatic; HIV infection;
KW   anti-HIV; immune disorder; gene ds.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   91..834

```
Db   1008  TCTTCATTTTTCTCCTTATAAGATGGACACCATATCTGACTCATACAAATTTTATAAT  1067

Qy   1514  TAATTTTCAAATCATCATAGAATTGATTTGAGCAGGGTCTTCCAGAATGCTTCCTTG  1573

Db   1068  TTATGAAAATTATTGGCATGGCTTTGGTGTGGAGGTAGGGACTCAAAGAATATTTCTTTC  1127

Qy   1574  TTCCTATCCATAGATAGCCATTCCCCACCTTTATCAGAGTGGTAATGAAAACTGTGCAATTGT  1633

Db   1128  TTACTATTTATACCAGCCATCTGCACCTTCACCAGACCGATAATAAAAGCATGCTTTCTT  1187

Qy   1634  GCCAAAGACCCTTCTGAAGAGATGTCTGAAT------CATGCGGCCGAGTTTTTACACA  1687

Db   1188  G---GGATCTCTTCTAAAGATGAAGATAGCTGCATTACCAGTATCCAGTGGATTTCTTTCTG  1244

Qy   1688  CAGCTCTTCCTTTATAAATAAATCCTTCCCATTCTCCCCTGTAGA---------------  1735

Db   1245  TGGCAGGTCTCATCATACTGGGAAGTCTCAAAGACTGCATGCAAGA  1304

Qy   1736  ------GTACAGAAACAAAAATACCCTGATGATTCAGCAAGAAGAAAGTCTTTTTACTTA  1788

Db   1305  AAATTATATACTTAAGCAAAAGTCTTGATCATTCGGGAATGAAATTCTCTCTGAGTA  1364

Qy   1789  --GCAAATGTGCCTGCTTCGATCAGTTCGCTTGTGACATTAAGCTAAGCTGGGTTTGGGTTTTTG  1846

Db   1365  TGGCAGTATGCCTGCTTCGATGAATTCACTTTAATAGT-AGCTGTGATGATATTTTG  1423

Qy   1847  GTTGGATTTGGGGCGTTCTTCACTTCTTTGTCTATATTTCCTTACCTTTATCAGTTT  1906

Db   1424  GTTGAATTTAAGCCATTTATTTCTCCTTATGTCTACATTTCTTCATCACTTAATAATTT  1483

Qy   1907  GTATTCGAGCTTCCTGCTTTGGGATTCTGCAATTCTCTCTCCCACTGACAGGATCAACTC  1966

Db   1484  -TATTTTAACTTCCTGCCTCATTCTTACACTTTTTTCTCTCTCTACTTATACGATATATACTT  1542

Qy   1967  AATGACATAAAGTAGTTCAAACATCCATTGCTTCTCACATGTT---TTATCCATAAAAGTT  2023

Db   1543  CAAGGAGCTGAACAGTTACAATACCCAGTCTTTCTTACCTGTTGTCTTGCCTGTGTGGGGTC  1602

Qy   2024  ACTCATCTGATTTTATTT--AAAATAGTGAACATCTACTTGATATCAGACCGAGGACCA  2081

Db   1603  ACTCACTCCAATTTTATTTTAAAATTACTGAATATAATACTTCATCGCCCAGCCC-TGCACCA  1661

Qy   2082  TCCTCCATTGGAGAATATGAAG-ATATTGTCACTGGCAGAAAGCAGGTGTGTGCCATT  2139

Db   1662  GGCTCTGTGAGACGACAAGAAGGATTATTGCCCTCTTGTAGGAGAAAACAGATGTGCATAT  1721

Qy   2140  AAT------TGATAAGGATACCACGAAGCATCATCATGCC---AGTTATGGAACACAGTGCTG  2190

Db   1722  CATTAATTATGATGGGGTATGAATAGAGTTACAATGACAAAAGCTATGAATAAAGTGCTA  1781

Qy   2191  AAAGGATCATAGACAGAGGGTGGTAAAATCTGATCCCAGTAGAATAAACTTCAGTGTAC  2248

Db   1782  CAAGGAACAGAAGCAGAGCAGGGATAATAAATC--CATTTCAGTGAGGGAATCTATTTCAATGTAC  1838
```

```
FT                /*tag= a
FT                /product= "cell division regulating protein 27"
XX
PN   CN1493590-A.
XX
PD   05-MAY-2004.
XX
PF   30-OCT-2002; 2002CN-00137726.
XX
PR   30-OCT-2002; 2002CN-00137726.
XX
PA   (SHAN-) SHANGHAI BIODOOR GENE DEV CO LTD.
XX
PI   Mao Y, Xie Y;
XX
DR   WPI; 2004-500615/48.
DR   P-PSDB; ADV69201.
XX
PT   New polypeptide-human cell disintegrate regulatory protein 27 for
PT   treating cancer, human immunodeficiency virus infection, and immunopathy.
XX
PS   Claim 6; SEQ ID NO 1; 25pp; Chinese.
XX
CC   The invention relates to human cell division regulating protein 27, the
CC   polynucleotide encoding it, preparing the polypeptide by DNA recombinant
CC   techniques, application of the polypeptide in treating diseases, such as
CC   cancer, human immunodeficiency virus (HIV) infection, immunopathy and the
CC   antagonist of the polypeptide and its medical action and application of
CC   the polynucleotide. The present sequence is that of the human cell
CC   division regulating protein 27 encoding DNA of the invention.
XX
SQ   Sequence 840 BP; 267 A; 178 C; 164 G; 231 T; 0 U; 0 Other;

Query Match      19.4%;     Score 614.8;  DB 13;  Length 840;
Best Local Similarity   83.6%;     Pred. No. 1.3e-150;
Matches  697;  Conservative    0;  Mismatches 137;  Indels    0;  Gaps    0;

Qy    440 TATGACAATTTTAGAGCAAAACTGGCTTCCTGTTGCGATGCCATTCAAGACTTCGTGGTT 499
          |||| ||||||||||||||||||||| |||||| | |||||||||| ||||||| ||||||
Db      1 TATGCAAATTTTAGAGCAAACTGCTTCCTGCTGTGATGCGTGTTCAAAACTTTGTGTT 60

Qy    500 TCCCAGAACAACACTCCAGTGGGGACTAACATGAGCTACGAGGTGGAAAGCAAGAAACAC 559
          |||| ||||| |||||||||||||||||||||||| ||||||||||||| |||||||||||
Db     61 TCTCAGAATAACACTCCAGTTGGGACTAATATGAGTTACGAGGTGGAAAGCAAAAAAGAA 120

Qy    560 ATCCCCATTCGAGAGAACATTTTCCACATGTTTCCAGTGTCCGCAGCCTTTTGTGGACTAT 619
          ||||| ||| |||||| ||||| |||| ||||||||||||| || |||||||||||||||||
Db    121 ATCCCAATTAAGAGAAGAACATTTTCATATGTTTCCAGTGTCCCAGCCTTTGTGGGACTAC 180

Qy    620 CCCTATAACCAGTGTGCAGTGGTTGGTAATGGGGGAATTCTCAACAAGTCTCTCTGCGGA 679
          |||||| |||||||||||||||||||||||||| ||| |||||||||| || ||||||||
Db    181 CCTTATAATCAGTGTGCAGTGGTTCGGAAATGGGGAATTCTGAATAAGTCTCTCTGTGGA 240

Qy    680 GCAGAAATTGATAAATCTGACTTCGTCTTCAGTGTAACCTCCCCCAATCACAGGAGAGC 739
          ||| |||||| |||| |||||||||||||||| |||||||||| |||||||||||||
Db    241 ACTGAAATAGATAAATCCGACTTCGTTTTTAGGTGTAACCTACCCCCAACCACAGAGAT 300

Qy    740 GCTAGTAAAGATGTTGGAAGCAAACAAATCTTGTGACTGTCAATCCCAGCATTATAACC 799
          ||||||||||||||| || |||| ||||||| || |||||||| |||||| |||| |||
Db    301 GTTAGTAAAGATGTTGGCAGTAAAACAAATCTTGTGACTATAAATCCAAGCATCATAACT 360

Qy    800 CTGAAGTACCAGAATTCAAGGAGGAAGAAGCACAGTTTTTGGAGGACATCTCCACCTAT 859
          ||||| || |||||| || ||||||||||||||| ||| || ||| |||||||||||||
Db    361 ACTGAAATATGGGAACTAAAGGAAAAAAAGCCCTATTTCTGGAGGACATTGCAACCTAT 420

Qy    860 GGAGATGCATTCCTCCTGCCAGCATTTCCATCGGGCCAACACAGGCATCTCTTTT 919
          |||||||| |||||||||||||||| || |||||| |||||||||||||||||| |||
Db    421 GGAGATGCATTTTTCTTCTGCCAGCATTTCCTTCAGGGCCAACACGGGTACCTCTTTC 480

Qy    920 AAAGTCTACCAAACACTCAAAGATCAAAAATGAGGCAAAAGTTCTTCCTTCCATCCC 979
          ||||| ||||| || |||| || |||||||||||||||||||| |||||||||||
Db    481 AAAGTATACTACACGCTCGAAGAGTTCAAAGCAAGAAGACAAAAGTTCATTTTTCCATCCC 540
```

```
RESULT 5
ABV78252
ID   ABV78252 standard; DNA; 1071 BP.
XX
AC   ABV78252;
XX
DT   15-NOV-2002  (first entry)
XX
DE   Human GD3 synthase DNA SEQ ID NO 136.
XX
KW   RNA inhibition; dsRNA1; gene expression inhibitor; oncogene; cytostatic;
KW   virucide; protozoacide; gene; ds
XX
OS   Homo sapiens.
XX
PN   WO200255693-A2.
XX
PD   18-JUL-2002.
XX
PF   09-JAN-2002; 2002WO-EP000152.
XX
PR   09-JAN-2001; 2001DE-01000586.
PR   26-OCT-2001; 2001DE-01055280.
PR   29-NOV-2001; 2001DE-01058411.
PR   07-DEC-2001; 2001DE-01060151.
XX
PA   (RIBO-) RIBOPHARMA AG.
XX
PI   Kreutzer R,  Limmer S,  Rost S,  Hadwiger P;
XX
DR   WPI; 2002-590671/63.
XX
PT   Inhibiting expression of target gene, useful e.g. for inhibiting
PT   oncogenes, by administering double-stranded RNA complementary to the
PT   target and having an overhang.
XX
PS   Claim 10; Page 193-194; 203pp; German.
XX
CC   The invention relates to inhibiting expression of a target gene (I) in a
CC   cell by introducing an inhibitory RNA (dsRNA1) having a double-stranded
CC   structure of at most 49 consecutive bases. At least part of one strand
CC   (asl) of dsRNA1 is complementary to (I) and at least one end of dsRNA1
CC   has an overhang of 1-4 nucleotides. The method is used to inhibit the
CC   expression of a wide range of genes, e.g. oncogenes, cytokine genes etc.
CC   in humans, also genes in Plasmodium or in viruses or viroids that are
CC   pathogenic for humans, animals or plants. Introducing an overhang into
CC   dsRNA greatly increases effectiveness for inhibiting gene expression,
CC   both in vivo and in vitro and also increases stability and thus the
CC   effective concentration inside the cell. The present sequence is that of
CC   a gene related to the invention
XX
SQ   Sequence 1071 BP; 264 A; 277 C; 269 G; 261 T; 0 U; 0 Other;
```

```
Qy    980 AGTGTACCTGAGACACCTCGCTCTTTCTGGAGAACTAAAGGGGTGACTGCATACCGCTTG 1039
          |||||||||||||||||||||||| |||||||||||||||||||||| ||||||||||||
Db    541 AGTGTACCTGAAAGATCTGGCCCTTTCTGGAGAACTAAAGGTGGACTGCATACCGCTTG 600

Qy   1040 TCCACAGGCTTGATGATTGCAAGGTGCCTGTGGAACTGTGTGAAAACGTGAAGCTCTAC 1099
          |||| |||||||||||||||||| ||||||||||||||||||| |||| |||||||||||
Db    601 TCCACCGGCTTGATGATCACAAGGTTGCAGTGGAACTGTGTAAAAATGTGAAGCTGTAT 660

Qy   1100 GGATTCTGGCCTTCTCTAAGACATCGAAGACACCCCACTCAGTCACCACTACTATGAT 1159
          |||| || |||||||| |||||||||||||||||||||||||||||||||||| |||||
Db    661 GGATTCTGGCCCTTCTCTAAAACGTAGAAGACATACCTGTCAGCCATCACTATTATGAC 720

Qy   1160 AACATGTTACCTAAGCACATGGTTTCCACCAGATGCCTAAAGAATACAGCCAAATGCTCCAG 1219
          |||| || ||||||||| ||||| ||||||||||||||| |||||| ||||||||||| ||
Db    721 AACAAGCTACCTAAACATGGTTTCATCAGATGCCCAAAGAATACAGCCAGATCCTCCAA 780

Qy   1220 CTCCATATGAGGAGAATCCTCAA CTGCAATTCAGCAAATGTGAAACGGCTTAA 1273
          || |||||||||||||||||||| ||||||||||||| |||||||| |||||
Db    781 CTTCACATGAAAGGAATCCTCA  ACTGCAATTTAGCAAAGTGAAGTGCCTAA 834
```

KW   protozoacide; gene expression; antisense; tumour; infection; Plasmodium;
KW   virus; viroid; anti-GFP; human; HIV; human immunodeficiency virus;
KW   Hepatitis C virus; human papilloma virus; gene; ds.
OS   Homo sapiens.
XX
PN   DE10100588-A1.
XX
PD   18-JUL-2002.
XX
PF   09-JAN-2001; 2001DE-01000588.
XX
PR   09-JAN-2001; 2001DE-01000588.
XX
PA   (RIBO-) RIBOPHARMA AG.
XX
PI   Kreutzer R, Limmer S, Rost S, Hadwiger P;
XX
DR   WPI; 2002-683450/74.
XX
PT   Inhibiting expression of target genes, useful e.g. for treating tumors,
PT   by introducing into cells two double-stranded RNAs that are complementary
PT   to the target.
XX
PS   Claim 13; Page 91; 100pp; German.
XX
CC   The invention relates to inhibiting expression of a target gene in a cell
CC   by introducing at least two oligoribonucleotides (dsRNAI and II), both
CC   with a double-stranded (ds) structure of at most 49 sequential nucleotide
CC   pairs. At least part of one strand (S1, S2) of the ds structures in each
CC   of dsRNAI and II are complementary to regions in the target gene. The
CC   method uses antisense inhibition of gene expression using double stranded
CC   RNA inhibition (RNAi). The method is particularly used to treat tumours
CC   or infections, especially by Plasmodium or viruses/viroids (pathogenic on
CC   humans, animals or plants). The method provides more effective inhibition
CC   of expression than known method using a single dsRNA, even at very low
CC   concentrations. When dsRNA has at least one unpaired nucleotide at the
CC   end, stability (and thus effective concentration in the cell) is improved
CC   and efficiency can be increased further by pretreating the cells with
CC   interferon. The present sequence is that of a target DNA of the invention
XX
SQ   Sequence 1071 BP; 264 A; 277 C; 269 G; 261 T; 0 U; 0 Other;

Query Match   7.1%;   Score 224.4; DB 6;   Length 1071;
Best Local Similarity   55.3%;   Pred. No. 6.4e-48;
Matches 457; Conservative   0; Mismatches 366;   Indels   3; Gaps   1;

Qy   449 TTTAGAGCAAAACTGGCTTCCTGTTGCGATGCCATTCAAGACTTCGTGGTTCCCAGAAC 508
           ||||| |||| |||||| |||||||||| ||||||| |||||||||||
Db   232 TTCAGGAAACAAATGAAGACTGCTGCGACCCTGCCATCTCTTTGCTATGACTAAAATG 291

Qy   509 AACACTCCAGTGGGGACTAACATGAGCTACGAGGTGAAAGCAAGAAACACATCCCCATT 568
           || ||||||| ||||||| ||||| ||||||||||| ||||
Db   292 AATTCCCCTATGGGGAAGAGCATGGTATGGATGACGGGGAGTTTTTATACTCATTCACCATT 351

Qy   569 CGAGAGAACATTTTCCACATGTGTTCCAGTGTCGCCAGCCTTTGTGGACTATCCCTATAAC 628
           | ||||||| ||||||||||||||||| |||| ---TCCAGCTGCCATTGAAG
Db   352 GACAATTCAACTTACTTACTCTCTTCCCACAGGCAACCCCAT---TCCAGCTGCCATTGAAG 408

Qy   629 CAGTGTGCAGTGGTTGGTAATGGGGAATTCTCAACAAGTCTCTCTGCCGGAGCAGAAATT 688
           ||||||||| | |||| ||||| |||||| ||||||| || | ||||||||
Db   409 AAATGCGCGGCGTGTCGTGGGAAATGTGGGATTCTGAAGAGAGTGCGTGTGCCCGTCAAATA 468

Qy   689 GATAAATCTGACTTGCTTCAGGTGAACCTCCCCCAATCACAGGGAGCGCGCTAGTAAA 748
           ||||| ||||||||||| |||||||||| || || |||| ||||||
Qy   469 GATGAAGCAAATTTTGTCATGCATGCAATCTCCTTGTCAAGTGAATACACTAAG 528

Qy   749 GATGTTGGAAGCAAAACAAATCTTGTGACTGTCAATCCCAGCATTATAACCTGAAGTAC 808
           ||||||||||| ||||| |||| ||||||||| |||||||||||||
Db   529 GATGTTGGATCCAAAAGTCAGTAGTGACAGCAGTAGCCAATAATTCGGCAAAGGTTT 588

Qy   809 CAGAATTTGAAGGAGGAGAAAGCACAGTTTTTGGAGGACAATCTCCACCTATGGAGATGCA 868
           ||||| |||||||||||| || || ||||||||||||||||
Db   589 CAGAACCTTCGTGGTCCAGAAAGACATTTGTGACAACAGCAATCTATAACCACAGT 648

Qy   869 TTCCTCCCCTGCCAGACATTTCCTATCGGGACCAACACAGGCATCTCTTTAAAGTCTAC 928
           |||| ||| |||||||||
Db   649 TACATCTACATGCCTGCCTTTTCTATGAAGACAGGAACAGAGCCATCTTTGAGGGTTTAT 708

Qy   929 CAAACACTCAAAGAGTCAAAAATGAGGCAAAAGTTCTCTTTCCATCCCAGGTACCTG 988
           ||||| || ||
Db   709 TATACACTGTCGACAGTGTGGTGCCCAATCAAACAGTGCGTTTGCCAACCCAACTTTCTG 768

Qy   989 AGACACCTCGCTCTTTTCTGGAGAACTAAAGGGGTGACTGCATACCGCTTGTCCACAGGC 1048
           |||| |||||||
Db   769 CGTAGCATTGGAAAGTTCTGGAAAAAGTAGGAGGAATCCATGCCAAGCGGCCTGTCCACAGA 828

Qy   1049 TTGATGATTGCAAGTGTCGCTGTGGGAACCTGTGAAAACGTGAAGCTCACGGATTCTGG 1108
           |||||||||| |||||||||||| |||||||||| |||||||||
Db   829 CTTTTTTCGGTGAGCGCAGCTCTGGGTCTCTGTTCTGTGAAGAGGTGGCCATCTATGGCTTCTCTGG 888

Qy   1109 CCTTTCTCTAAGACTATCGAAGACACCCCACTCAGTCACCACTACTATGATAACATGTTA 1168
           || |||||||||||| |||||||| |||||||| |||||||||||
Db   889 CCCTTCTCTGTGAATATGCATGGATCATGAGCAGCGCCACCAGCCACCTACTATGAGACGTCTTA 948

Qy   1169 CCTAAGCATGGTTTCCACCAGATGCCTAAAGAATACAGCCAAATGCTCCAGCTCCATATG 1228
           |||||| |||||||||||| |||||||||| |||||||||
Db   949 CCCTTTTCTGGCTTCCATGCCATGGCCCGGAATTTCTCCAACTCTGGTATCTTCATAAA 1008

Qy   1229 AGAGGAATCCTCAACTGCAATTCAGCAAATGTGAAACGGCTTAAC 1274
           ||||||| ||||| |||||||
Db   1009 ATCGGTGCACTGAGAATGCAGCTGGACCCATGTGAAGATACCTCAC 1054

RESULT 6
ABZ35828
ID   ABZ35828 standard; DNA; 1071 BP.
XX
AC   ABZ35828;
XX
DT   07-FEB-2003 (first entry)
XX
DE   Human GD3 synthase polynucleotide SEQ ID NO 136.
XX
KW   Double stranded RNA; dsRNA; RNAi; RNA inhibition; cytostatic; virucide;

CC   target genes, particularly oncogenes, cytokine genes, Id (not defined)
CC   protein genes; developmental or prion genes, or genes expresed in
CC   pathogenic organisms (particularly plasmodia) or in viruses or viroids
CC   (pathogenic in humans, animals or plants). Treating the cells with
CC   interferon greatly increases the extent to which dsRNA can inhibit
CC   expression of the target genes, and the effect is even greater when dsRNA
CC   are modified to increase their stability. ABX09936-ABX10075 represent
XX   gene fragments used to illustrate the method of the invention
SQ   Sequence 1071 BP; 264 A; 277 C; 269 G; 261 T; 0 U; 0 Other;

Query Match   7.1%; Score 224.4; DB 6; Length 1071;
Best Local Similarity   55.3%; Prd. No. 6.4e-48;
Matches 457; Conservative   0; Mismatches 366; Indels   3; Gaps   1;

```
Qy   449 TTTAGAGCAAAACTGGCTTCCCT...TTGGATGCCAATCGAGCTTCGTGGTGGTTTCCCAGAAC 508
             ||||  |||  | |   |||       ||| |||     | | |||  |||  |  ||   
Db   232 TTCAGGAAACAAATGGAAGACTC...TGCGACCCTGCCATCTCTTTGCTATGACTAAAATG 291

Qy   509 AACACTCCAGTGGGGACTAACA...AGTACGAGGTGGAAACAAGAGAAACATCCCATT 568
             ||    ||||| ||| | |       ||| ||  |||   |    | | | ||| ||||
Db   292 AATTCCCCATGGGGAAGAGCA...TTGGTATGACGGGGAGTTTTTATACTCATTCACCATT 351

Qy   569 CGAGAAGAACATTTCCACATGT...TCCAGTGTCGCAGCCTTTTGTGGACTATCCCTATAAC 628
             | ||| | ||  |  | | ||     |||| |    ||  ||  |||||| ||  | ||||
Db   352 GACAATTCAACTTACTCTCTCT...CCCACAGGCAACCCCAT---TCCAGCTGCCATTGAAG 408

Qy   629 CAGTGTGCTAGTGGTTGGTAATG...GGGAATTCTCTCGCGAGCAGAAATT 688
             |||| |    || ||| || |||    ||||  | | || |||||  |||
Db   409 AAATGCGCGGTGGTGGGAAATG...GTGGGGATTCTGAAGAGAGTGGCCGTCAAATA 468

Qy   689 GATAAATCTGACTTCGTCTTCA...GTGTAACCTCCCCCAATCACAGGGAGCGCTAGTAAA 748
             || |||  ||   | ||| |||     ||  ||||| |     |  |||  || |  |||
Db   469 GATGAAGCAAATTTTGTCATGC...ATGCAATCTCCCTCCTTTGTCAAGTGAATACACTAAG 528

Qy   749 GATGTTGGAAGCAAAACAAATC...TGTGACTGTCAATCCCAGCATTATAACCCTGAAGTAC 808
             ||||||||| ||||||  |||      ||| | ||| |||||||| | || |  |||| |  
Db   529 GATGTTGGATCCAAAAGTCAGT...AGTGACAGCTAATCCCAGCATAATTCGGCAAAGGTTT 588

Qy   809 CAGAATTTGAAGAGGAAGAAAG...ACAGTTTTTGGAGGACATCTCCACCTATGGAGGATGCA 868
             ||||| | |  | | ||| |||    |||  |||| |||||| |    | | | ||||||
Db   589 CAGAACCTTCGTGTGGTCCAGAA...GACATTTGTGGACAACATGAAAAATCTATAACCACAGT 648

Qy   869 TTCCTCCCTGCCAGCATTT...CTATCGGGACCAACACAGGAGCCATCTCTTTAAGTCTAC 928
             | |||| || || |  ||       |  || | |||| ||| ||||||| ||| ||| |||
Db   649 TACATCTACATGCCGCCTTTT...TATGAAGAACAGGAACAGAGCCCATCTTTGAGGGTTTAT 708

Qy   929 CAAACACTCAAAGAGTCAAAAA...AGGCAAAAGGTTCTCTTCTTCCATCCCAGGTACCTG 988
             ||||||||| || || |||||    || ||||||| || ||| | | | |||  | ||||
Db   709 TATACACTGTCAGATGTTGGTG...CAATCAAACAGTGCTGTTTGCCAACCCCAACTTTCTG 768

Qy   989 AGACACCTCGCTCTCTTTTCTGA...CTAAAGGGTGACTACCGGCATACCGGCTTGTCCACAGGC 1048
             ||||||| || || | |||  ||    | ||| | |  |  || ||| || || |||||||| 
Db   769 CGTAGCATTGGAAAGTTCTGA...AAGTGGAGGAATCCATGGCCAAGCGCCTGTCCACAGGA 828

Qy  1049 TTGATGATTGCAAGTGTCGCTG...GGAACTGTGTGAAAACGTGAAGCTTCACGGATTCTGG 1108
             | |  ||||||||| || | |     ||| ||||||| | |||||| | ||| || || |||
Db   829 CTTTTTCGTGGTGAGCGGAGCTC...GGGTCTCTGTGAAGAGAGTGGCCATCTATGGCTTCGG 888

Qy  1109 CCTTTCTCTAAGACTATCGAAGC...CACCCCACTAGTCACCACTACTACTGATAACATGTTA 1168
             || || || |  |  |   | ||    ||||| |  ||||||||||||||| || |||  ||
Db   889 CCCTTCTCTGTGAATATGCATG...CAGCCCATCAGCACCACCACTACTGACAAGGTCTTA 948

Qy  1169 CCTAAGCATGGTTTCCACCAGA...GCCTAAAGAATACAGCCAAATGCTCCAGCTCCATATG 1228
             || || ||||| ||||| | |     || || ||| |  || ||| || ||||| ||||
Db   949 CCCTTTCTGGGCTTCCATGCCA...GCCCGAGGAATTTCTCCAACTCTGGTATCTTCATAAA 1008

Qy  1229 AGAGGAATCCTCAAACTGCAAT...CAGCAAATGTGAAACGGCTTAAC 1274
             ||| || | | ||| || ||     | |||||||||| || |  ||
Db  1009 ATCGGTGCACTGAGAATGCAGC...CGGACCCCATGTGAAGATACCTCAC 1054
```

RESULT 8

```
Db   589 CAGAACCTTCGTGTGGTCCAGCAGCATTTGTGGACAACATGAAAAATCTATAACCACAGT 648
Qy   869 TTCCTCCCTGCCAGCATTTCCTATCGGGTCCTCGTGTGGTGCTG...ACCAACACAGGAGCCATCTCTTTAAGTCTAC 928
Db   649 TACATCTACATGCCGCCTTTTCCGGAGGAATTCTCCAACTCTGGTATCTTCATAAA... 708
Qy   929 CAAACACTCAAAGAGTCAAAAA...GAGGCAAAAGGTTCTCTTCTTCCATCCCAGGTACCTG 988
Db   709 TATACACTGTCAGATGTTGGTG...CAATCAAACAGTGCTGTTTGCCAACCCCAACTTTCTG 768

Qy   989 AGACACCTCGCTCTCTTTTCTGGA...GCATACCGGCTTGTCCACAGGC 1048
Db   769 CGTAGCATTGGAAAGTTCTGGA...AAGTAGGGGAATCCATGGCCAAGCGCCTGTCCACAGGA 828

Qy  1049 TTGATGATTGCAAGTGTCGCTG...GGAACTGTGTGAAAACGTGAAGCTTCACGGATTCTGG 1108
Db   829 CTTTTTCGTGGTGAGCGGAGCTC...GGGTCTCTGTGAAGAGAGTGGCCATCTATGGCTTCGG 888

Qy  1109 CCTTTCTCTAAGACTATCGAAGC...CACCCCACTAGTCACCACTACTACTGATAACATGTTA 1168
Db   889 CCCTTCTCTGTGAATATGCATG...CAGCCCATCAGCACCACCACTACTGACAAGGTCTTA 948

Qy  1169 CCTAAGCATGGTTTCCACCAGA...GCCTAAAGAATACAGCCAAATGCTCCAGCTCCATATG 1228
Db   949 CCCTTTCTGGGCTTCCATGCCA...GCCCGAGGAATTTCTCCAACTCTGGTATCTTCATAAA 1008

Qy  1229 AGAGGAATCCTCAAACTGCAAT...GAAACGGCTTAAC 1274
Db  1009 ATCGGTGCACTGAGAATGCAGC...CGGACCCCATGTGAAGATACCTCAC 1054
```

RESULT 7
ABX10071
ID   ABX10071 standard; DNA; 1071 BP.
XX
AC   ABX10071;
XX
DT   23-JAN-2003 (first entry)
XX
DE   Human GD3 synthase DNA fragment SEQ ID 136.
XX
KW   Oligoribonucleotide; interferon; oncogene; cytokine; Id; developmental;
KW   prion; inhibition; human; ds.
XX
OS   Homo sapiens.
XX
PN   DE10100587-C1.
XX
PD   21-NOV-2002.
XX
PF   09-JAN-2001; 2001DE-01000587.
PR   09-JAN-2001; 2001DE-01000587.
XX
PA   (RIBO-) RIBOPHARMA AG.
XX
PI   Kreutzer R, Limmer S, Rost S, Hadwiger P;
XX
DR   WPI; 2002-742209/81.
XX
PT   Inhibiting expression of target genes, e.g. oncogenes, in cells, by
PT   introduction of complementary double-stranded oligoribonucleotide, after
PT   treating the cell with interferon.
XX
PS   Disclosure; Page 96; 98pp; German.
XX
CC   This invention describes a novel method for inhibiting expression of a
CC   target gene by introducing into the cell that contains the target gene at
CC   least one oligoribonucleotide (dsRNAi) that has a double-stranded (ds)
CC   structure of not more than 49 consecutive nucleotides (nt), where at
CC   least a segment of one strand of the ds structure is complementary with
CC   the target gene and the cells are treated with interferon before
CC   introduction of dsRNAi. The method is used to inhibit expression of

ABL91793
ID   ABL91793 standard; DNA; 1071 BP.
XX
AC   ABL91793;
XX
DT   28-MAY-2002 (first entry)
XX
DE   Human polynucleotide SEQ ID NO 136.
XX
KW   Human; HIV; HCV; gene expression; oligoribonucleotide; tumour; pathogen;
KW   Plasmodium; virus; viroid; cytokine; prion; antisense oligonucleotide;
KW   cytostatic; virucide; protozoacide; antibacterial; ds.
XX
OS   Homo sapiens.
XX
PN   DE10100586-C1.
XX
PD   11-APR-2002.
XX
PF   09-JAN-2001; 2001DE-01000586.
PR   09-JAN-2001; 2001DE-01000586.
XX
PA   (RIBO-) RIBOPHARMA AG.
XX
PI   Kreutzer R, Limmer S, Rost S, Hadwiger P;
XX
DR   WPI; 2002-270454/32.
XX
PT   Inhibiting gene expression in cells, useful for e.g. treating tumours, by
PT   introducing double-stranded complementary oligoRNA having unpaired
PT   terminal bases.
XX
PS   Claim 13; Page 97; 104pp; German.
XX
CC   The invention relates to a method for inhibiting expression of a target
CC   gene (ABL91658-ABL91797) in a cell by introducing at least one
CC   oligoribonucleotide that has a double-stranded structure consisting of at
CC   most 49 sequential nucleotide pairs, with at least part of one strand
CC   complementary with the target gene and has at least one end a single-
CC   stranded segment of 1-4 nt. The method provides oligoribonucleotides for
CC   antisense inhibition of gene expression useful e.g. for treating tumours
CC   but the oligoribonucleotides may also be directed against genes present
CC   in pathogens (e.g. Plasmodium or viruses/viroids, pathogenic on humans,
CC   animals or plants) or against cytokine, Id, developmental or prion genes.
CC   The method provides more effective inhibition of gene expression than use
CC   of known oligonucleotides, probably because the unpaired overhang
CC   increases stability and thus intracellular concentration
XX
SQ   Sequence 1071 BP; 264 A; 277 C; 269 G; 261 T; 0 U; 0 Other;

Query Match   7.1%;   Score 224.4; DB 6; Length 1071;
Best Local Similarity 55.3%;   Pred. No. 6.4e-48;
Matches 457; Conservative   0; Mismatches 366; Indels   3; Gaps   1;

Qy   449 TTTAGAGCAAAACTGGCTTCCTGTTGGGATGCCATTCAAGACTTCGTTGGTTTCCCAGAAC 508
            || |||| ||||||||| |||| |||||| | |||| | ||||| ||||||||||||||
Db   232 TTCAGGAAACAAATGGAAGACTGCTGGGACACCCTGCCCATCTCTTTGCTATGACTAAAATG 291

Qy   509 AACACTTCCAGTGGGGACTAACATGAGCTACGAGGTACGAAAGCAAGAACACATCCCCATT 568
            ||| || || || ||||| || |||| ||| || ||| | |||||| || |||| ||||
Db   292 AATTCCCCTATGGGGAAGAGCATGTGGTATGACGGGGAGTTTTATACTCATTCACCATT 351

Qy   569 CGAGAGAACATTTTCCACATGTTTCCAGTGTCGCAGCCTTTGTGACTATCCCTATAAC 628
            |||||||||| | |||||| ||||| ||| |||| |||| | | |||||| ||||
Db   352 GACAATTCAACTTACTCTCTCTTCCACAGCAACCCAT---TCCAGCTGCCATTGAAG 408

Qy   629 CAGTGTGCAGTGGTTGGTAATGGGGAATTCAACAAGTCTCTCTGCGGAGCAGAAATT 688
            |||| ||||| | |||| |||||| ||||| |||||||||||| ||| | ||||||
Db   409 AAATGCCGCGGTGTGGGGAAATGTGGGGATTCTGAAGAAGAGAGTGGCTGGCCGCCGTCAAATA 468

Qy   689 GATAAATCTGACTTCGTCTTCAGGTGTAACCTCCCCCAATCACAGGGAGCGCTAGTAAA 748
            ||| | ||| | || | | ||| |||| | |||| || || ||| |||
Db   ...

Db   469 GATGAAGCAAATTTGTCATGCGCATGCAATCTCCCTCCTTGTCAAGTGAATACATACTAAG 528
Qy   749 GATGTTGGAAGCAAACAAATCTTGTGACTGTCAATCCCAGCATTATAACCCTGAAGTAC 808
              ||||||||||| |||| | |||| || || |||||| || ||| |||||||||
Db   529 GAGTGTTGGATCCAAAAGTCAGTTGACAGCTAATCCCAGCATAATTCGGCAAAGGTTT 588

Qy   809 CAGTTTTGAGGAGAAGAAGGCACAGTTTCTGGACAACATGAATCTATAACCACAGT 868
             ||| | ||||||||||| || || || | || || ||| | | | |||
Db   589 CAGAACCTTCGTGGTCCAGAACACATTTGTGACAACATGAAAAATCTATAACCACAGT 648

Qy   869 TTCCTCCTCCTGCCAGCATTTCGTATCGGGCCAACACAGGCATCTCTTTTAAAGTCTAC 928
            ||| | | | | | | ||| ||| |||| | || || | | || || |||
Db   649 TACATCTACATGCCTGCCTTTTCCATGAAGACAGCAACAGCGAACAGAGCCATCTTTGAGGGTTTAT 708

Qy   929 CAAACACTCAAAGAGTCAAAAATGAGGCAAAAGTTCTTCTCTTCCATCCCAGTGACCTG 988
             ||| |||||| || | |||||||| | | |||||||| | || || || ||||||
Db   709 TATACACTGTCAGATGTTGGTGCCAATCAAACAGTGCTGTTTGCCAACCCCAACTTTCTG 768

Qy   989 AGACACCTCGCTCTCTTTTCGGACAACTAAAGGGGTGACTGCATACCGCTTGTCCACAGGC 1048
            || |||| ||| ||||| ||||| || ||| || | | | ||||| ||||||||
Db   769 CGTAGCATTGGAAAGTTCTGGAAAGAGTAGTAGAGGAATCCATGCCCAGGCGCTGTCCACAGGA 828

Qy   1049 TTGATGATTGCAAGTGTCGCTGTGGACTGTGTGAAAACGTGAAGCTTCACGGATTCTGG 1108
             |||| |||||||||| | ||| || | | | |||| ||| | || | ||||||||
Db   829 CTTTTTCTTGGTGAGCGACGCTCGGGTCTTCGTGTGAAGAGTGGCCATCTATGGCTTCTGG 888

Qy   1109 CCTTTCTCTAAGACTATCGAAGACACCCCACTCAGTCAGTCACCACTACTATGATAACATGTTA 1174
            |||||| || ||| | || | || | || | ||| ||||| | || |||| | |||||
Db   889 CCCTTCTCTGTGAATATGCATGAGCAGCCCATCAGCCCAAATGCTCCAGTCTCCATATG 948

Qy   1169 CCTAAGCATGGTTTCCACCAGATGCTAAAGAATACAGCCAAATGCTCCAGTCCCATATG 1228
            ||||| |||| | || || || | ||| | | | || || || |||| || ||||
Db   949 CCCTTTTCTGGCTTCCATGCCACCCCGAGGATTTCTCCAACTCGGTATCTTCCATAAA 1008

Qy   1229 AGAGGAATCCTCAAACTGCAATTCAGCAAAATGTGAAACGGCTAAC 1274
             | | ||| | |||| || | ||| ||| ||| | | |||
Db   1009 ATCGGTGCACTGAGAATGCAGCCTGACCCATGTGAAGATACCTCAC 1054

RESULT 9
AAQ77831
ID   AAQ77831 standard; cDNA to mRNA; 2117 BP.
XX
AC   AAQ77831;
XX
DT   25-MAR-2003 (revised)
DT   26-JUN-1995 (first entry)
XX
DE   Human alpha-2,8-sialyltransferase gene.
XX
KW   Human; alpha-2,8-sialyltransferase; sialyltransferase; melanoma; E.coli;
KW   pUC119; expression vector; pAMoRC3Sc; PCR; amplify; primer; sugar chain;
KW   physiological; ganglioside; ds.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   483..1553
FT   /tag= a
FT   /product= "alpha-2,8-sialyltransferase"
XX
PN   WO9423020-A1.
XX
PD   13-OCT-1994.
XX
PF   28-MAR-1994; 94WO-JP000495.
XX
PR   29-MAR-1993; 93JP-00069988.
XX
PA   (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI   Sasaki K, Kurata K, Hanai N, Nishi T;
XX

Qy 1169 CCTAAGCATGGTTTCCACCAGATGCCTAAAGAATACAGCGAAATGCTCCAGCTCCATATG 1228
          |||| ||| |||| |||| ||| ||| |||
Db 1431 CCCTTTCTGGCTTCCATGCCATGCCCGGAGGAATTTCTCCAACTCTGGTATCTTCATAAA 1490

Qy 1229 AGAGGAATCCTCAAACTGCAATCAGCAAATGTGAAACGGCTAAC 1274
          | | ||| ||| ||| |||| ||| ||||| | ||
Db 1491 ATCGGTGCACTGAGAATGCAGCCGGACCCATGTGAAGATACCTCAC 1536

RESULT 10
ADR24998
ID   ADR24998 standard; DNA; 2117 BP.
XX
AC   ADR24998;
XX
DT   21-OCT-2004 (first entry)
XX
DE   Breast cancer prognosis marker #359.
XX
KW   ds; breast cancer; prognosis; gene expression; diagnosis.
XX
OS   Homo sapiens.
XX
PN   WO2004065545-A2.
XX
PD   05-AUG-2004.
XX
PF   15-JAN-2004; 2004WO-US001100.
XX
PR   15-JAN-2003; 2003US-00342887.
XX
PA   (ROSE-) ROSETTA INPHARMATICS LLC
PA   (NECA-) NETHERLANDS CANCER INST
XX
PI   Van't Veer LJ, He Y;
XX
DR   WPI; 2004-593473/57.
XX
PT   Classifying a breast cancer patient according to prognosis comprises
PT   determining the similarity between the level of expression of each of
PT   five genes in a cell sample taken from patient, to control levels.
XX
PS   Disclosure; SEQ ID NO 859; 226pp; English.
XX
CC   The invention relates to a method of classifying a breast cancer patient
CC   according to prognosis by determining the similarity between the level of
CC   expression of each of five genes for which markers are listed in the
CC   specification, in a cell sample taken from the breast cancer patient, to
CC   control levels of expression for each respective five genes to obtain a
CC   patient similarity value. The methods are useful for classifying a breast
CC   cancer patient according to prognosis. Kits and computer program products
CC   are useful for data analysis using the diagnostic, prognostic and
CC   statistical methods of the invention. This sequence corresponds to a
CC   marker used in the method of the invention.
XX
SQ   Sequence 2117 BP; 554 A; 502 C; 571 G; 490 T; 0 U; 0 Other;

Query Match   7.1%; Score 224.4; DB 13; Length 2117;
Best Local Similarity   55.3%; Pred. No. 9.2e-48;
Matches 457; Conservative   0; Mismatches 366; Indels   3; Gaps   1;

Qy   449 TTTAGAGCAAAACTGGCTTCCTGTTGCGATGCCATTCAAGACTTCGTGGTTTCCCAGAAC 508
          || || ||| | || | | ||||| | ||||||||| | | | || || |
Db   714 TTCAGGAAACAAATGGAAGACTGCTGCGACCCTGCCCATCTCTTTGCTATGACTAAAATG 773

Qy   509 AACACTTCCAGTGGGGACTAACATGAGCTACGAGGTGAAAGCAAGAAGAAACACATCCCATT 568
          ||| || ||||| || || ||| |||| |||| | | ||| || ||| |||
Db   774 AATTCCCTATGGGAAGACATGTGGTATGACGGGGAGTTTTATACTCATTCACCATT 833

Qy   569 CGAGAGAACATTTTCCACATGTTCCAGTGTCGCAGCCCTTTGTGGGACTATCCCTATAAC 628
          | ||| |||| || | |||| ||| |||| | ||| ||| |||| |||||| |
Db   834 GACAATTCAACTTACTTCTCTCTCCCACAGGCAACCCCAT---TCCAGCTGCCATTGAAG 890

Best Local Similarity   55.3%; Pred. No. 9.2e-48;
Matches 457; Conservative   0; Mismatches 366; Indels   3; Gaps   1;

DR   WPI; 1994-333182/41.
DR   P-PSDB; AAR63225.
XX
PT   New alpha-2,8,-sialyl transferase - useful for producing physiologically
PT   active sugar chains.
XX
PS   Claim 3; Page 63-66; 102pp; Japanese.
XX
CC   The nucleotide sequence of the novel gene encoding the human alpha-2,8-
CC   sialyltransferase. The gene encodes a protein of 356 a.a. The gene
CC   (called WP1) was isolated from the human melanoma cell line WM266-4 and
CC   cloned into the E.coli plasmid pUC119 to create pUC119.WP1. The gene was
CC   inserted into the expression vector pAMoPRC3Sc to produce the plasmid
CC   pAMoPRSAWP1. The missing N-terminal 19 a.a. were inserted into this
CC   vector by PCR amplification using the primers AAQ77835-6. Alpha-2,8-
CC   sialyltransferase is useful in the production of physiologically active
CC   sugar chains e.g. in the conversion of ganglioside GM3 to GD3. (Updated
CC   on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence 2117 BP; 554 A; 502 C; 571 G; 490 T; 0 U; 0 Other;

Query Match   7.1%; Score 224.4; DB 2; Length 2117;
Best Local Similarity   55.3%; Pred. No. 9.2e-48;
Matches 457; Conservative   0; Mismatches 366; Indels   3; Gaps   1;

Qy   449 TTTAGAGCAAAACTGGCTTCCTGTTGCGATGCCATTCAAGACTTCGTGGTTTCCCAGAAC 508
          || || ||| | || | | ||||| | ||||||||| | | | || || |
Db   714 TTCAGGAAACAAATGGAAGACTGCTGCGACCCTGCCCATCTCTTTGCTATGACTAAAATG 773

Qy   509 AACACTTCCAGTGGGGACTAACATGAGCTACGAGGTGAAAGCAAGAAGAAACACATCCCATT 568
          ||| || ||||| || || ||| |||| |||| | | ||| || ||| |||
Db   774 AATTCCCTATGGGAAGACATGTGGTATGACGGGGAGTTTTATACTCATTCACCATT 833

Qy   569 CGAGAGAACATTTTCCACATGTTCCAGTGTCGCAGCCCTTTGTGGGACTATCCCTATAAC 628
          | ||| |||| || | |||| ||| |||| | ||| ||| |||| |||||| |
Db   834 GACAATTCAACTTACTTCTCTCTCCCACAGGCAACCCCAT---TCCAGCTGCCATTGAAG 890

Qy   629 CAGTGTGCAGTGGTTGGTAATGGGAATTCTCAACAAGTCTCTCTGCGGAGCAGAAATT 688
          ||| | | |||||||||| || |||| |||||| |||||| ||||| ||||
Db   891 AAATGGCGCGGTGGTGGGAAATGTGGGATTCTGAAGAGAGTGGCTGTGGCCGTCAAATA 950

Qy   689 GATAAATCTGACTTCGTTCTTCAGGTGATGCCGATGCAATCTCCCCTCCCCAATCACAGGCCTAGTAAA 748
          ||| || | || | || | | ||||||| || ||| ||| | || || |||||
Db   951 GATGAAGCAAATTTGTCATGCGATGCAATCTCCCCTCCTTGTCAAGTGAATACACTAAG 1010

Qy   749 GATGTTGGAAGCAAACAAATCTTGTGACTGTCAATCCAGCATTATAAACCCTGAAGTAC 808
          ||||||||| |||| | | | ||||| || |||||||||| || | ||| ||| |
Db 1011 GATGTTGGATCCAAAAGTCAAAGTCAGTTAGTGACAGCTAATCCCAGCATAATTCGGCAAAGGTTT 1070

Qy   809 CAGAATTTGAAGGAGGAGAAGCACAGTTTTGGAGGACAATCTCCACCTATGGAGATGCA 868
          |||| | | || | | |||| | |||| |||| | || | |||||||||||
Db 1071 CAGAACCTTCTGTGGTCCAGAAAGACATTTGTGGACACAATGAAAATCTATAACCACAGT 1130

Qy   869 TTCCTCCCTGCCCAGCATTTCCTATCGGCCAACAACAGCAGGCCATCTCTTTAAAGTCTAC 928
          | | || | ||||||| ||| | | ||||||| | || ||||| ||| |||
Db 1131 TACATCTACACATGCCTGCCTTTTCTATGAAGACAGAACAGAGCCATCTTTGAGGGTTTAT 1190

Qy   929 CAAACTCAAAGAGTCAAAAATGAGGCAAAAGGTTCTTCCATCCCCAGGTACCTG 988
          | |||| | | | |||| |||| | ||| | | | || | || ||| ||||
Db 1191 TATACACTGTCAGATGTTGGTGCCAATCAAACAGTGCTGTTTGCCAACCAACTTTCTG 1250

Qy   989 AGACACCTCGCTCCTTTTCTGGAGAACTAACGGCTGCTGCATACCGGCTTGTCCACAGGC 1048
          ||| | | ||| || || | || | || || ||||| | | | ||||||||
Db 1251 CGTAGCATTGGAAAGTTCTGAAAAGTAGAGAATCCATGCCAAGCGCCTGTCCACAGGA 1310

Qy 1049 TTGATGAATTGCAAGTGCTGCTGGAACTGTGTGAAAACGTGAAGCTTCACGGATTCTGG 1108
          ||| | | | ||| |||| || ||||||||||| ||| ||| | ||||| |||||
Db 1311 CTTTTCTGGTGAGCGACGCTCGGGTCTCTGTGAAGAGGTGGCCATCTATGGCTTCTGG 1370

Qy 1109 CCTTTTCCTAAGCTTCTGAAGACATCGAAGACACCTCACCACTACTATGATAACATGTTA 1168
          || |||| | || || | | | | | || | | | |||||||||||||| || |||
Db 1371 CCCTTTCTGTGAATATGCAATGGAGCAATGCAGCCCATCAGCCACCACTACTATGAACAACGTCTTA 1430

```
Qy    629  CAGTGTGCAGTGGTGGTAATGGGGAATTCTCAACAAGTCTCTGCGGAGCAGAAATT  688
                |||||  ||||||| |  ||||||| |||||||  |||||  ||| |||||
Db    891  AAATGCGCGGTGGTGGGGAAATGGTGGGATTCTGAAGAAGAGTGGCGTGGCCGTCAAATA  950

Qy    689  GATAAATCTGACTTCGTCTTCAGGTGTAACCTCCCCCAATCACAGGAGGCGCTAGTAAA  748
                |||||||  |  ||| ||||  ||||||| ||  |||||  || ||| |||||||
Db    951  GATGAAGCAAATTTGTCATGCGATGCAATCTCCCTCCTTGTCAGTGAATACACTAAG  1010

Qy    749  GATGTTGGAAGCAAAACAAATCTTGTGACTCAATCCCAGCATTATAACCCTGAAGTAC  808
                |||||| |  |||||  ||||||  |  |  || |||| |  || || ||||
Db    1011 GATGTTGGATCCAAAAGTCAGTTAGTGACAGCAGCTAATCCAGCATAATTCGGCAAAGGTTT  1070

Qy    809  CAGAATTTGAAGGAGGAAGAAGCACAGTTTTTGGAGGACATCTCCACCTATGGAGATGCA  868
                ||||| ||||  ||| |  |||| ||| |  ||| |||| || |||  ||
Db    1071 CAGAACCTTCGTGTGGTCCGAGAAGACATTTGTGACAACATGAAAAATCTATAACCACAGT  1130

Qy    869  TTCCTCCCTCCGCCAGCATTTCCTATCGGGCCAACACAGGCATCTCTTTAAAGTCTAC  928
                || ||| || ||  |  ||||  |||| ||| ||  |  ||  |||  || |||
Db    1131 TACATCTACATGCCGCCTTTCTATGAAGACAGGAACAGAGCCATCTTTGAGGGTTTAT  1190

Qy    929  CAAACACTCAAAGAGTCAAAAATGAGGCAAAAGTTCTCTTCCATCCCAGGTACCTG  988
                ||||| ||| ||||| |  |  | | |||||||| | ||| ||| |||| |||
Db    1191 TATACACTGTCAGATGTGGTGCCAATCAAACAGTGCGTGTTGCCAACCCCAACTTTCTG  1250

Qy    989  AGACACCCGCTCTTTCTCGGAGAACTAAAGGGTGACTGCATACCGTTGTCCACAGGC  1048
                |||||||  |||||||||||||||  ||  ||  ||  ||  ||||||||||||
Db    1251 CGTAGCATTGGAAAGTTCTGGAAAGTAGAGGAAAGTAGGTAGCAATTCATGCCAAGCGGCCTGTCCACAGGA  1310

Qy    1049 TTGATGATTGCAAGTGTCGTGTGGGACTGTGTGAAAACGTGAAGCTCTACGGATTCTGG  1108
                ||||| || |||||  ||||   | ||  || |||  ||||  |||||||||||||
Db    1311 CTTTTTCGGTGAGCGCAGCTCGGGTCTTCGTGGTCCATCTATACTACTACTATGGCTTCTGG  1370

Qy    1109 CCTTTCTCTAAGACTATCGAAGACACCCCACTGAGTCACCACTACTACTATGATAACATGTTA  1168
                ||| |||||||  |  || ||| ||  || ||| | ||| | | || |||| || ||||
Db    1371 CCCTTCTCTGTGAATATGCATGAGCAGAGCCATCAGCCACCACTACTACGAGTCGTCTTA  1430

Qy    1169 CCTAAGCATGGTTTCCACCAGATGCCTAAAGATACAGCCAAATGCTCCAGCTCCATATG  1228
                |||||||||| ||| ||  |  ||| || || |||  ||||||||  || | ||||
Db    1431 CCCTTTTCTGGCTTCCATGCCGATGGCCCGGAGAATTTCTCCAACTTGCTGTATCTTCATAAA  1490

Qy    1229 AGAGGAATCCTCAAACTGCAATTCAGCAAGTGAAAATGTGAAACGGCTTAAC  1274
                 | || |  | ||||  ||| |||| ||  ||||||||  ||||||
Db    1491 ATCGGTCACTCGACTGAGAATGCAGCTGGACCCATGTGAGATACCTCAC  1536
```

RESULT 11
AAT03694
ID   AAT03694   standard;   cDNA to mRNA; 1704 BP.
XX
AC   AAT03694;
XX
DT   07-AUG-1996   (first entry)
XX
DE   Human alpha 2,8-sialyltransferase cDNA.
XX
KW   Human melanoma; SK-Mel-28; ganglioside 3-synthase; tumour antigen;
KW   detection; Northern blotting; glycosylation; ds.
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   149..1174
FT   /tag= a
FT   /product= "alpha-2,8-sialyltransferase"
FT   unsure   721..766
FT   /tag= b
FT   /note= "the sequence of bases 721-766 is missing from the
FT   specification. Generic codons coding for the amino acid
FT   sequence VDNMKIY- NHSYIYMP have been inserted by the
FT   indexer"
XX
PN   JP07327678-A.
XX

```
PD    19-DEC-1995.
XX
PF    03-JUN-1994;    94JP-00145450.
XX
PR    03-JUN-1994;    94JP-00145450.
XX
PA    (SEGK ) SEIKAGAKU KOGYO CO LTD.
XX
DR    WPI; 1996-072336/08.
DR    P-PSDB; AAR88326.
XX
PT    Alpha 2,8-sialyl:transferase cDNA - useful for sensitive detection of
PT    tumour antigen by Northern blotting.
XX
PS    Claim 1; Page 5-8; 10pp; Japanese.
XX
CC    A cDNA clone coding for alpha 2,8-sialyltransferase was isolated from a
CC    human melanoma SK-Mel-28 cDNA library and sequenced. The insert codes for
CC    a protein having ganglioside 3-synthase activity which is useful as a
CC    biochemical reagent. The cDNA is useful for detecting tumour antigen by
CC    Northern blot analysis
XX
SQ    Sequence 1704 BP; 452 A; 377 C; 434 G; 426 T; 0 U; 15 Other;

Query Match             7.0%;  Score 222.2; DB 2; Length 1704;
Best Local Similarity  54.4%;  Prcd. No. 3.1e-47;
Matches 449;  Conservative   8;  Mismatches 366;  Indels   3;  Gaps  1;

Qy    449  TTTAGAGCAAACTGGCTTCCTTGCGATGCATTCAAGAGACTTCGTGGTTTCCAGAAC  508
                ||  ||||  ||  | ||||  ||  ||||||  || |||  |||||  ||||||
Db    335  TTCAGAAACAAATGGAGAGACTCTGCGACCCTGCCATCTCTTTGCTATGACTAAAATG  394

Qy    509  AACACTCCAGTGGGACTAACACAGAGCTACGAGGTGGAAAGCAAGAAAACACATCCCATT  568
                |||| ||  ||| | |  ||||  ||||||||| | || | ||| |||||  | ||
Db    395  AATTCCCCTATGGGAAGAGCAGTGGTATGACGGGAGAGTTTTATACTCATTCACCATT  454

Qy    569  CGAGAGAACAGATTTCCACATGTCGCAGTGTCGCAGCCCTTTGTGGACTATCCCTATAAC  628
                || ||  |  | | ||||||   ||  |||  | |      ||  |||  ||||||
Db    455  GACAATTCAACTTACTCTCTCT---TCCAGTGCCAACCCCAT---TCCAGCTGCCATTGAAG  511

Qy    629  CAGTGTGCAGTGGTTGGTAATGCGGGAATTCTCAACAGTCCTCTCGGAGCAGCAGAAATT  688
                |||||||  |  | ||| ||  |  |||||||| |  ||  |  | | || |  | ||
Db    512  AAATGCGCGGTGGTGGGGAAATGGTGGGATTCTGAAGAAGAGTGGCGTGGCCGTCAAATA  571

Qy    689  GATAAATCTGACTTCGTCTTCA     GTGTAACCTCCCCCAATCACAGGAGGCGCTAGTAAA  748
                |||| |   |  |||   |  |        |||||  || |||  | | ||| || ||||
Db    572  GATGAAGCAAATTTGTCATGCGATGCAATCTCCCTCCTTGTCAGTGAATACACTAAG  631

Qy    749  GATGTTGGAAGCAAAACAAATCTTGTGACTCAATCCCAGCATTATAACCCTGAAGTAC  808
                ||||| || |  |||| ||||  |  |  || |||| |  || || ||||
Db    632  GATGTTGGATCCAAAAGTCAGTTAGTGACAGCAGCTAATCCAGCATAATTCGGCAAAGGTT  691

Qy    809  CAGAATTTGAAGGAGGAAGAAGCACAGTTTTTGGAGGACATCTCCACCTATGGAGATGCA  868
                |||| ||  | | | |    |  ||  ||  || | |:::  |  | ||||| |||
Db    692  CAGAACCTTCGTGTGGTCCAGAAGACATTNGAYAVAYATGAAARATHTAYAAYCAYWSN  751

Qy    869  TTCCTCCCTCCGCCAGCATTTCCTATCGGGCCAACACAGGCATCTCTTTTAAGTCTAC  928
                || || | |  || | || |:|:       |   | | ||   |  ||  |||| |||
Db    752  TAYAYHTAYATGCCGCCTTTTTAT       TAYGAAGACAGGAACAGAGCCATCTTTGAGGGTTTAT  811

Qy    929  CAAACACTCAAAGAGTCAAAAAT GAGGCAAAAGTTCTCTTCCATCCCAGGTACCTG  988
                ||  ||| |||||  |   |||     |||||  ||| ||| |||| |  || |||
Db    812  TATACACTGTCAGATGTTGGTG     CAATCAAACAGTGCTGTTGCCAACCCCAACTTTCTG  871

Qy    989  AGACACCCGCTCTTTTCTCGGAG     ACTAAAGGGGTGACTGCATACCGCTTGTCCACAGGC  1048
                |||||||  ||||| ||||||      ||  |||| | |||  || || |  ||||||
Db    872  CGTAGCATTGGAAAGTTCTGGA     AAGTAGAGGAATTCCATGGACCGCCTGTCCACAGGA  931

Qy    1049 TTGATGATTGCAAGTGTCGCTG     TGGAACGTGAAAAACGTGAAAGCTCTACGGATTCTGG  1108
                ||| || | |  |  ||  |       || | | || ||| || || ||  |||||
Db    932  CTTTTTCGGTGAGCGCAGCTC     TGGGTCTTCGTGGTCTCGTGAAGAGGTGGCCTTCTG  991

Qy    1109 CCTTTCTCTAAGCTATCGAAGACACCCCACTGAGTCACCACTACTACTATGATAACATGTTA  1168
```

```
                                          Matches 442; Conservative  0; Mismatches 370; Indels  3; Gaps  1;

Db   992 CCCTTCTCTGTGAATATGCATGAGCAGCCATCAGCCACCACTACTATGACACGTCTTA 1051     Qy  449 TTTAGAGCAAACTGGCTTCCTGTGCGATGCCATTCAAGACTTCGTGGTTCCCAGAAC  508
                  || |||||| ||| ||    || ||||||||||||||||| |||| |||                 |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Qy  1169 CCTAAGCATGGTTCCACCAGATGCCATGAGAATACAGCCAAATGCTCCAGCTCCATATG 1228     Db  280 TTCAGGAGACAAATGGAAGACTCTGCAACCCCGCCCATCTCTTTGCAATGACGAAGGTG  339
                  || |||||||| |||||||| |||||                                          |||||||||||||||||||||||||||||||||||||||||||
Db  1052 CCCTTTTCTGGCTTCCATGCCATGCCCGGAGGAATTTCTCCAACTCTGTATCTTCATAAA 1111     Qy  509 AACACTCCAGTGGGACTAACAGAGCTACGAGTGGAAAGCAAGAAACACATCCCCATT  568
                                                                                      |||||||
Qy  1229 AGAGGAATCCTCAAACTGCAATTCAGCAAATGTGAAACGGCTTAAC 1274              Db  340 AACTCCCCATGGGAAGAGCCGTGGTATGACGGGGAGTTCTATACTCGCTCCACCATC  399
                  |  ||| ||||| |||| |||| |||||| || |||||                             |||
Db  1112 ATCGGTGCACTGAGAATGCAGCTCGGCGGACCCATGTGAAGATACCTCAC 1157          Qy  569 CGAGAGAACACATTTCCACATGTTTCAGTGTCGCAGCCCTTTGTGGACTATCCCTATAAC  628
                                                                                      ||||||  |||||
                                                                 Db  400 GACAACTCCACGTACTCCCTCT---TCCAGCTGCCATTGAAG  456

RESULT 12                                                          Qy  629 CAGTGTGCAGTGGTTGGTAATGCGGGAATTCTCAACAAGTCTCTCTGCCGACAGCAGAAATT  688
ADB58122                                                                 ||  |||| |||  |||||  |||||||
ID  ADB58122 standard; DNA; 1223 BP.                             Db  457 AAATGTGCGGTGGTGGGAAAACGTGGAATTCTGAAGATGAGTGGCTGTGGCCGTCAAATA  516
XX
AC  ADB58122;                                                     Qy  689 GATAAATCTGACTTCGTCTTCAGGTGTAACCTCCCCCAATCACAGGAGCGCTAGTAAA  748
XX                                                                      |||||||||  ||||  |||||  || |||||
DT  04-DEC-2003  (first entry)                                   Db  517 GATGAAGCAAATTTTGTCATGCGGTGTAACCTTCCTCCCCTTGTCAAGTGAATACACCAGA  576
XX
DE  Toxicity-related gene, SEQ ID 3148.                          Qy  749 GATGTTGGAAGCAAAACAAATCTGGTGACTGTCAATCCCAGCATTATAACCCTGAAGTAC  808
XX                                                                      |||||  ||  ||||| ||||||||  ||  |||
KW  Toxic; toxin; gene expression profile; hepatotoxicity; liver;  Db  577 GATGTGGGTTCCAAGACTCAGTGGTGCAGCTAACCCCAGCATAATTCGCCAGAGATTT  636
KW  drug screening; toxicity assay; ds.
XX                                                                Qy  809 CAGAATTTGAAGGAGAAGAAAGCACAGTTTTTGGAGGACGCATCTCCACCTATGGAGGATGCA  868
OS  Unidentified.                                                        |||| ||||||||||||||
XX                                                                Db  637 GAAAACCTGCTGTGGTCCAGAAGAAGTTGTGGACAACAGAGATCTATAACCACCAGT  696
PN  WO200306624-A2.
XX                                                                Qy  869 TTCCTCCTCCTGCCAGCATTTCCTATCGGGCCAACACAGGCATCTCTTTTTAAAGTCTAC  928
PD  07-AUG-2003.                                                         ||||||  |||||||||||||
XX                                                                Db  697 TACATATACATGCTGCCCTTTCTCATGAAGACGGGAACAGAGCCGTCTCTCCGTGTGTAC  756
PF  31-JAN-2003; 2003WO-US003194.
XX                                                                Qy  929 CAAACACTCAAAGAGTCAAAAATCAGGCAAAAGTTCTTCTTTCCATCCCAGGTACCTG  988
PR  31-JAN-2002; 2002US-00060087.                                       |  |||  |||||||||
PR  15-MAR-2002; 2002US-0364045P.                                Db  757 TACACCCTGAAAGACGCTGGAGCAATCAAACAGTGTCTTTTGTCAACCCAACTTTCTT  816
PR  15-MAR-2002; 2002US-0364055P.
PR  30-DEC-2002; 2002US-0436643P.                                Qy  989 AGACACCTCGCTCTTTTCTGGACTAAAGGGGTGACTGCATACCGCTTGTCCACAGGC  1048
XX                                                                      ||| ||||||||||||||
PA  (GENE-) GENE LOGIC INC.                                      Db  817 CGAAACATTGGAAAGTTCTGGACGGGTAGGGGGGATCCATGCCAAGCGCCTGTCTACAGGA  876
XX
PI  Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;    Qy 1049 TTGATGATTGCAAGTGTGCGCTGTGGAACTGTGTGAAAACGTGAAGCTTCACGGATTCTGG  1108
XX                                                                      ||||||||| |||||||
DR  WPI; 2003-689530/65.                                        Db  877 CTCTTTTTGGCTTGGGTAGTGCAGCCTTGGGCCTTTGTGAGGAGGTGTCCATCTACGATGTTCTCGG  936
XX
PT  Predicting a toxic effect of a compound, useful in identifying toxicity   Qy 1109 CCTTTCTTCTTAAGACTATCGAAGCACCCCAGTCAGTCACCACTACTATGATAACATGTTA  1168
PT  markers in liver tissues or cells for drug screening and toxicity assays,        ||||||||
PT  comprises preparing gene expression profile of tissue or cells exposed to  Db  937 CCCTTCTCCGTGAACATGCAAGGGAACCTATCAGCCACCATTACTATGACAACGTCTTG  996
PT  the compound.
XX                                                                Qy 1169 CCTAAGCATGGTTCCACCAGACGCCTAAAGAATACAGCCAAATGCTCCAGCTCCAGCTCCATATG  1228
PS  Claim 1; SEQ ID NO 3148; 1156pp; English.                          |||||||||||||
XX                                                                Db  997 CCCTTCTCAGGCTTCCATGCCACGGCCTGAGGAGTTTCTTCAACTCTGGTATCTTCATAAA  1056
CC  The present invention relates to a method for predicting a toxic effect
CC  of a compound. The method comprises preparing a gene expression profile   Qy 1229 AGAGGAATCCTCAAACTGCAATTCAGCAAATGTGA  1263
CC  of a tissue or cell sample exposed to the compound, and comparing the            ||| ||
CC  gene expression profile to a database comprising SEQ ID 1-4925, where     Db 1057 ATGGGTGCACTGAGGATGCAACCGACCCCATGTGA  1091
CC  differential expression of the gene indicates at least one toxic effect.
CC  The method is useful for predicting hepatotoxicity or the progression of a
CC  compound, predicting hepatotoxicity or the progression of a toxic effect   RESULT 13
CC  of a compound, identifying an agent that modulates the onset or           AAH34115
CC  progression of a toxic response, predicting the cellular pathways that a   ID  AAH34115 standard; cDNA; 1511 BP.
CC  compound modulates in a cell, and identifying an agent that modulates at   XX
CC  least one activity of a protein. The method and compositions of the       AC  AAH34115;
CC  present invention using a database of genes having liver toxin-induced     XX
CC  differential expression, are useful in identifying toxicity markers in    DT  03-SEP-2001  (first entry)
CC  liver tissues or cells for drug screening and toxicity assays. Note: The   XX
CC  sequence data for this patent did not form part of the printed            DE  Human colon cancer antigen encoding cDNA SEQ ID NO:1197.
CC  specification, but was obtained in electronic format directly from WIPO    XX
CC  at ftp.wipo.int/pub/published_pct_sequences.                              KW  Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX                                                                KW  colorectal carcinoma, chromosome 12; ss.
SQ  Sequence 1223 BP; 281 A; 322 C; 331 G; 289 T; 0 U; 0 Other;   XX
                                                                 OS  Homo sapiens.
Query Match             6.5%;   Score 207; DB 10; Length 1223;
Best Local Similarity  54.2%;   Pred. No. 2.6e-43;
```

```
                    557  AACTTCTGCGTAR--TTGGAAASTTCTGAAAASTAGAGAWTCCATGCCAAGCGGCTG  614
Db

Qy                 1040  TCCACAGGCTTGATGATTGCAAGTGTCGCTGTGGAACTGTGTGAAAAGTGAAGCTCTAC 1099
                         ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db                  615  TCCACAGGACTTTTTCTGGTGAGCGCACTTGSGGNTCTCTGTGAAGAGGTGGCCATCTAT  674

Qy                 1100  GGATTCTGCCTTTCTCTAAGACTATGAAGACACCCCACTCAGTGCACTACTATGAT 1159
                         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db                  675  GGCTTCTGGCCCTTCTGTGACATGCATGAGCAGCCCATGCACCACTACTATGAC  734

Qy                 1160  AACATGTTACCTAAGCATGTTTCACCAGATGCCTAAAGAATACAGCCAAATGTCTCCAG 1219
                         ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db                  735  AACGTCTTACCCTTTTCTGGCTTTCCATGCCATGCCGAGGAATTTCTCCAACTCTGGTAT  794

Qy                 1220  CTCCATATGAGAGAATCCTCAAACTGCAATTCAGCAAATGTGAAACGGCTTAAC 1274
                         ||||| ||||| ||||||||||||||||||||||
Db                  795  CTTCATAAAATCGGTGCACTGACAATGCAGCTGACCCATGTGAAGATACCTCAC  849
```

RESULT 14
ABA56318
ID   ABA56318 standard; DNA; 448 BP.
XX
AC   ABA56318;
XX
DT   01-FEB-2002 (first entry)
XX
DE   Human foetal liver single exon nucleic acid probe #4623.
XX
KW   Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157277-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000669.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Eank DR;
XX
DR   WPI; 2001-483447/52.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human fetal liver.
XX
PS   Claim 1; SEQ ID NO 4623; 639pp 1 Sequence Listing; English.
XX
CC   The invention relates to a single exon nucleic acid probe for measuring
CC   human gene expression in a sample derived from human foetal liver. The
CC   single exon nucleic acid probes may be used for predicting, measuring and
CC   displaying gene expression in samples derived from human fetal liver. The
CC   present sequence is a single exon nucleic acid probe of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 448 BP; 114 A; 112 C; 98 G; 124 T; 0 U; 0 Other;

Query Match   4.2%; Score 131.4; DB 4; Length 448;
Best Local Similarity 57.4%; Pred. No. 1.1e-23;
Matches 237; Conservative   0; Mismatches 176; Indels   0; Gaps   0;
```
```

XX
PN   WO200122920-A2.
XX
PD   05-APR-2001.
XX
PF   28-SEP-2000; 2000WO-US026524.
XX
PR   29-SEP-1999; 99US-0157137P.
PR   03-NOV-1999; 99US-0163280P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR   WPI; 2001-235357/24.
DR   P-PSDB; AAG74710.
XX
PT   Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT   useful for preventing, diagnosing and/or treating colorectal cancers.
XX
PS   Claim 1; Page 2986; 9803pp; English.
XX
CC   AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC   cancer-associated nucleic acid molecules (N) and proteins (P), where the
CC   proteins are collectively known as colon cancer antigens. The colon
CC   cancer antigens have cytostatic activity and can be used in gene therapy
CC   and vaccine production. N and P may be used in the prevention, diagnosis
CC   and treatment of diseases associated with inappropriate P expression. For
CC   example, N and P may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of P by expressing inactive proteins or to
CC   supplement the patients own production of P. Additionally, N may be used
CC   to produce the colon cancer-associated Ps, by inserting the nucleic acids
CC   into a host cell and culturing the cell to express the proteins. N and P
CC   can be used in the prevention, diagnosis and treatment of colorectal
CC   carcinomas and cancers. AAH37196 to AAH37204 and AAB77789 represent
CC   sequences used in the exemplification of the present invention. N.B.
CC   Pages 666 to 682 and page 7053 of the sequence listing were missing at
CC   time of publication, meaning no sequences are present for SEQ ID NO:1027
CC   to 1052, 7921 and 7922
XX
SQ   Sequence 1511 BP; 447 A; 304 C; 332 G; 420 T; 0 U; 8 Other;

Query Match   5.5%; Score 175.4; DB 4; Length 1511;
Best Local Similarity 56.8%; Pred. No. 5.8e-35;
Matches 372; Conservative   3; Mismatches 276; Indels   4; Gaps   3;

```
Qy    620  CCCTATAACCAGTGTGCAGTGGTTGGTAATGGGAGAATTCTCAACAGTCTCTCTGCGGA  679
           ||||                     |||      |||||||||||||||||||||||||
Db    199  CCATTGAAGAAATGCGCGGCGGTGGTGGGAAATGGTGGGATTCTGAAGAGATGGCTGTGGC  258

Qy    680  GCAGAAATTGATAAATCGACTTCGTCTTCAGGTGAACTCCCCCCCAATCACAGGGAGC  739
           ||||                      |||    ||||||||||||| ||||||||||||
Db    259  G-TCAAATAGATGAAGCAAATTTTGTCATGCGATGCAATCTCCCTCCCTTGTCAAGTGAA  317

Qy    740  GCTAGTAAAGATGTTGGAAGCAAAACAAAATCTTGTGACTGACTATTATACC  799
           |||||||||||||||||||| ||||||||| |||: |||
Db    318  TACACTAAGGATGTNGGATCCAAAAGTCAGTAGTAGTGACAGCTAATCCCAGCATAATTCGG  377

Qy    800  CTGAAGTAGTACCAGAATTGAAGGAGAAGAAGCACAGTTTTTGGAGACATCTCCACCTAT  859
           ||||||||||||||||||      |||||||| |||    :  :|||
Db    378  CAAAGGTTTCAGACCTTCAGTCGTCCAGAAGACAATTGTGACAACATGAAAATYTAT  437

Qy    860  GGAGATGCATTCCTCCCTGCCCAGCATTTCCTATCCGGCCAACACAGGCATCTCTTTT  919
           |||||||||||||||||||| ||||:| |||||| |||:||
Db    438  AACCACAGTTACATCTACATGCCTGCCTTTT-CTATGAAGACRGGAACAGAGCCATCTTG  496

Qy    920  AAAGTCTACCAAACACTCAAAGAGTCAAAAAATGAGGCAAAAGTTCTTCTTTCCATCCC  979
           |||||||||||||||||||| ||||  |||  ||||||||||||||||||| |||||
Db    497  AGGGTTTTATTATACACTGTCAGATGTTGGTGGTCCAATCAAACAGTGTGTTGCCAACCCC  556

Qy    980  AGGTACCTGAGACACCTCGCTCTTTTCGGAGAACTAAAGGGGTGACTGCATACCGCTTG 1039
           ||||||| :   ||     |||     ||||||   ||| ||
```

```
Qy  836 TTTTTGGAGGACATCCACCTATGGAGATGCATTCCTCCTGCCAGCATTTTCCTAT 895
         ||||  ||||                        |    | ||||  |||| |||||
Db   29 TTTGTGGACAACATGAAAATTTATAACCACAGTTACATCTACATGCCTGCCTTTTCTATG 88

Qy  896 CGGGCCAACACAGGCATCTCTTTAAAGTCTACCAAACACTCAAAGAGTCAAAAATGAGG 955
         |||| |     ||  ||||                ||  |||||     |||||
Db   89 AAGACAGGAACAGAGCCATCTTTTATTATACACTGTCAGATGTTGGTGCCAAT 148

Qy  956 CAAAAGGTTCTCTTCTTCCATCCCAGGTACCTGAGACACCTCGCTCTTTTCTGGAGAACT 1015
         ||||   |||      |||||       |||||     |||||||||  |||||||
Db  149 CAAACAGTGCGTGTTTGCCAACCCCAACTTTCTGCGTAGCATTGGAAAGTTCTGGAAAAGT 208

Qy 1016 AAAGGGGTGACTGCATACCGCTTGTCCACAGGCTTGATGATTGCAAGTGTCGCTGTGGAA 1075
         |   ||||| || || || |||||||||||||   |  ||     | |||  |||||||
Db  209 AGAGGAATCCATGCCAAGCGCCTGTCCACAGGACTTTTCTCGGTGAGCGCAGCTCTGGGT 268

Qy 1076 CTGTGTGAAAAACGTGAAGCTCTACGGATTCTGGCCTTTCTCTAAGACTATGCAAGACACC 1135
         || ||||||    ||| | ||||   |||||||| |||||||||| ||| ||||  |||||
Db  269 CTCTGTGAAGAGAGTGGCCATCTATGGCTTCTGCGCCCTTCTCTGTGAATATGCATGAGCAG 328

Qy 1136 CCACTCAGTCACCACTACTAATGATAACATGTTACCTAAGCAGCATGGTTTCCACCAGATGCCT 1195
          || || || ||||||||||| ||  ||     || || |  |  |  |||| |||  |||
Db  329 CCCATCAGCCACCACTACTATGACAACGTCTTACCCTTTCTTTCTGGCTTCCATCCATGCCC 388

Qy 1196 AAAGAATACAGCCAAATGCTCCAGCTCCATATGAGAGAATCCTCAAACTGCA 1248
          | |||| |||  ||||     |    ||| | |  |  ||  |||||||||
Db  389 GAGGAATTTCTCCAACTCTGGTATCTTCATAAAAATCGGTGCACTGAGAATGCA 441
```

RESULT 15
ABA45804
ID   ABA45804 standard; DNA; 448 BP.
XX
AC   ABA45804;
XX
DT   01-FEB-2002 (first entry)
XX
DE   Human breast cell single exon nucleic acid probe #4499.
XX
KW   Human; microarray; single exon probe; gene expression; breast; disease;
KW   cancer; 88.
XX
OS   Homo sapiens.
XX
PN   WO200157271-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000662.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00633366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR   WPI; 2001-496933/54.
XX
PT   New spatially-addressable set of single exon nucleic acid probes, useful
PT   for measuring gene expression in sample derived from human breast,
PT   comprises number of single exon nucleic acid probes.
XX
PS   Claim 1; SEQ ID NO 4499; 327pp + Sequence Listing; English.
XX
CC   The invention relates to a spatially-addressable set of single exon
CC   nucleic acid probes for measuring gene expression in a sample derived
CC   from human breast and BT 474 cells. The method involves contacting the
CC   probes with a collection of detectably labelled nucleic acids derived
CC   from mRNA of human breast, and then measuring the label bound to each
CC   probe of the microarray. The probes are useful for verifying the
CC   expression of regions of genomic DNA predicted to encode proteins. They
CC   are useful for gene discovery, and for determining predisposition and/or
CC   prognosing breast disease. Gene expression analysis is useful for
CC   assessing the toxicity of chemical agents on cells. The microarray of
CC   this invention presents a far greater diversity of probes for measuring
CC   gene expression, with far less than expressed sequence tag
CC   microarrays. The method is suitable for rapid production of functional
CC   information from genomic sequence. The present sequence is a single exon
CC   nucleic acid probe of the invention. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 448 BP; 114 A; 112 C; 98 G; 124 T; 0 U; 0 Other;

Query Match   4.2%; Score 131.4; DB 4; Length 448;
Best Local Similarity   57.4%; Pred. No. 1.1e-23;
Matches 237; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

```
Qy  836 TTTTTGGAGGACATCCACCTATGGAGATGCATTCCTCCTGCCAGCATTTTCCTAT 895
         ||||  ||||                        |    | ||||  |||| |||||
Db   29 TTTGTGGACAACATGAAAATTTATAACCACAGTTACATCTACATGCCTGCCTTTTCTATG 88

Qy  896 CGGGCCAACACAGGCATCTCTTTAAAGTCTACCAAACACTCAAAGAGTCAAAAATGAGG 955
         |||| |     ||  ||||                ||  |||||     |||||
Db   89 AAGACAGGAACAGAGCCATCTTTTATTATACACTGTCAGATGTTGGTGCCAAT 148

Qy  956 CAAAAGGTTCTCTTCTTCCATCCCAGGTACCTGAGACACCTCGCTCTTTTCTGGAGAACT 1015
         ||||   |||      |||||       |||||     |||||||||  |||||||
Db  149 CAAACAGTGCGTGTTTGCCAACCCCAACTTTCTGCGTAGCATTGGAAAGTTCTGGAAAAGT 208

Qy 1016 AAAGGGGTGACTGCATACCGCTTGTCCACAGGCTTGATGATTGCAAGTGTCGCTGTGGAA 1075
         |   ||||| || || || |||||||||||||   |  ||     | |||  |||||||
Db  209 AGAGGAATCCATGCCAAGCGCCTGTCCACAGGACTTTTCTCGGTGAGCGCAGCTCTGGGT 268

Qy 1076 CTGTGTGAAAAACGTGAAGCTCTACGGATTCTGGCCTTTCTCTAAGACTATGCAAGACACC 1135
         || ||||||    ||| | ||||   |||||||| |||||||||| ||| ||||  |||||
Db  269 CTCTGTGAAGAGAGTGGCCATCTATGGCTTCTGCGCCCTTCTCTGTGAATATGCATGAGCAG 328

Qy 1136 CCACTCAGTCACCACTACTAATGATAACATGTTACCTAAGCAGCATGGTTTCCACCAGATGCCT 1195
          || || || ||||||||||| ||  ||     || || |  |  |  |||| |||  |||
Db  329 CCCATCAGCCACCACTACTATGACAACGTCTTACCCTTTCTTTCTGGCTTCCATCCATGCCC 388

Qy 1196 AAAGAATACAGCCAAATGCTCCAGCTCCATATGAGAGAATCCTCAAACTGCA 1248
          | |||| |||  ||||     |    ||| | |  |  ||  |||||||||
Db  389 GAGGAATTTCTCCAACTCTGGTATCTTCATAAAAATCGGTGCACTGAGAATGCA 441
```

Search completed: May 31, 2006, 10:49:15
Job time : 1886 secs

```
                    GenCore version 5.1.8
          Copyright (c) 1993 - 2006  Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on:      May 31, 2006, 15:43:12 ; Search time 3701 Seconds
                                    (without alignments)
                            10511.390 Million cell updates/sec

Title:        US-10-501-930-2
Perfect score: 3166
Sequence:     1 cggagcggcgagtcggtgcc..........gctattagcaaaaaaaaa 3166

Scoring table: IDENTITY_NUC
              Gapop 10.0 , Gapext 1.0

Searched:     18892170 seqs, 6143817638 residues

Total number of hits satisfying chosen parameters:     37784340

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                Maximum Match 100%
                Listing first 45 summaries

Database :  Published Applications NA Main:*
            1: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US07_PUBCOMB.seq:*
            2: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US08_PUBCOMB.seq:*
            3: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
            4: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
            5: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
            6: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
            7: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
            8: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
            9: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
           10: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
           11: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
           12: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10G_PUBCOMB.seq:*
           13: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US11A_PUBCOMB.seq:*
           14: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US11B_PUBCOMB.seq:*
           15: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US11C_PUBCOMB.seq:*
           16: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US11D_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                            SUMMARIES
```

Result No.	Score	Query Match %	Match Length	DB	ID	Description
1	3166	100.0	3166	12	US-10-501-930-2	Sequence 2, Appli
2	930.2	29.4	1500	12	US-10-501-930-4	Sequence 4, Appli
3	278.2	8.8	810	12	US-10-301-480-562554	Sequence 562554,
c 4	278.2	8.8	810	6	US-10-301-480-1175963	Sequence 1175963,
c 5	254	8.0	672	7	US-10-027-632-44555	Sequence 44555, A
6	254	8.0	672	7	US-10-027-632-44555	Sequence 44555, A
7	224.4	7.1	1071	7	US-10-085-117-66	Sequence 66, Appl
8	224.4	7.1	1071	8	US-10-384-339C-136	Sequence 136, App
9	224.4	7.1	2117	7	US-10-172-118-859	Sequence 859, App
10	224.4	7.1	2117	7	US-10-085-117-65	Sequence 65, Appl
11	224.4	7.1	2117	7	US-10-430-325-1	Sequence 1, Appli
12	224.4	7.1	2117	8	US-10-342-887-859	Sequence 859, App
13	210	6.6	1026	7	US-10-085-117-63	Sequence 63, Appl
14	210	6.6	1388	7	US-10-085-117-62	Sequence 62, Appl
15	206.6	6.5	1223	16	US-11-136-527-2140	Sequence 2140, Ap
16	175.4	5.5	1511	6	US-10-106-698-1207	Sequence 1207, Ap
17	131.4	4.2	448	3	US-09-864-761-15072	Sequence 15072, A

18	130.8	4.1	428	3	US-09-864-761-21173	Sequence 21173, A
19	127.4	4.0	600	16	US-11-136-527-6236	Sequence 6236, Ap
20	112.2	3.5	150	3	US-09-867-550-1959	Sequence 1959, Ap
21	83.4	2.6	1080000	10	US-10-928-446A-1	Sequence 1, Appli
22	83.4	2.6	1080000	10	US-10-928-446A-181	Sequence 181, App
23	83.4	2.6	1080000	10	US-10-928-446A-183	Sequence 183, App
24	83.4	2.6	1080000	10	US-10-928-446A-185	Sequence 185, App
25	83.4	2.6	1080000	10	US-10-928-446A-187	Sequence 187, App
26	83.4	2.6	1080000	10	US-10-928-446A-189	Sequence 189, App
27	83.4	2.6	1080000	10	US-10-928-446A-191	Sequence 191, App
28	83.4	2.6	1080000	10	US-10-928-446A-193	Sequence 193, App
29	83.4	2.6	1080000	10	US-10-928-446A-195	Sequence 195, App
30	83.4	2.6	1080000	10	US-10-928-446A-197	Sequence 197, App
31	83.4	2.6	1080000	10	US-10-928-446A-199	Sequence 199, App
32	83.4	2.6	1080000	10	US-10-928-446A-201	Sequence 201, App
c 33	79.4	2.5	293	3	US-09-867-550-1829	Sequence 1829, Ap
34	75	2.4	472	3	US-09-918-995-29326	Sequence 29326, A
35	57.6	1.8	551	4	US-09-925-065A-748151	Sequence 748151,
36	57.6	1.8	551	5	US-09-925-065A-748152	Sequence 748152,
37	57.6	1.8	551	5	US-09-925-065A-748151	Sequence 748151,
38	57.6	1.8	551	5	US-09-925-065A-748152	Sequence 748152,
39	53.2	1.7	494	3	US-09-918-995-1105	Sequence 1105, Ap
c 40	52.6	1.7	542	7	US-10-029-386-13566	Sequence 13566, A
41	52.6	1.7	601	9	US-10-363-345A-34255	Sequence 34255, A
c 42	52.6	1.7	601	9	US-10-363-345A-34256	Sequence 34256, A
43	52.6	1.7	601	10	US-10-363-483A-34255	Sequence 34255, A
c 44	52.6	1.7	601	10	US-10-363-483A-34256	Sequence 34256, A
c 45	52.2	1.6	1416	7	US-10-311-455-2204	Sequence 2204, Ap

```
                            ALIGNMENTS

RESULT 1
US-10-501-930-2
; Sequence 2, Application US/1050193(
; Publication No. US2006005769 6A1
; GENERAL INFORMATION:
; APPLICANT:   TAKASHIMA, Shou
; APPLICANT:   TSUJIMOTO, Masafumi
; APPLICANT:   TSUJI, Shuichi
; TITLE OF INVENTION: GLYCOSYLATING ENZYME
; FILE REFERENCE: P25687
; CURRENT APPLICATION NUMBER: US/10,501,930
; CURRENT FILING DATE:  2004-07-29
; PRIOR APPLICATION NUMBER: PCT/JP03/00883
; PRIOR FILING DATE: 2003-01-30
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2
;   LENGTH: 3166
;   TYPE: DNA
;   ORGANISM: Mouse
US-10-501-930-2

Query Match           100.0%;  Score 3166;  DB 12;  Length 3166;
Best Local Similarity 100.0%;  Pred. No. 0;
Matches 3166;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy      1 CGGAGCGGCGAGTCGGTGCCCGGCTGCCTTCGCCCCGGCAGCTTTGGCGGCGAGG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CGGAGCGGCGAGTCGGTGCCCGGCTGCCTTCGCCCCGGCAGCTTTGGCGGCGAGG 60

Qy     61 ACGCCCGTGGCTCAGGATGAGATCGGGGGCACGCTGTTCGCCCTCATAGGCAGCCTGAT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ACGCCCGTGGCTCAGGATGAGATCGGGGGCACGCTGTTCGCCCTCATAGGCAGCCTGAT 120

Qy    121 GCTGCTCTGCTCCTCCTGCGTATGCCTCGGTGCCCAGCCGGCCGGCGGCGCTCCAGGCT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GCTGCTCTGCTCCTCCTGCGTATGCCTCGGTGCCCAGCCGGCCGGCGGCGCTCCAGGCT 180

Qy    181 GTTGATGAGGAAGCAGAGAGACCACCAGTGTACCTCAGCTGCACTGAAGACACTCTG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   181  GTTGATGAGGAGGCAGAGAGACCAGTGTACCTGCACTGAAGAGACACTCTG  240
Qy   241  GAGCCCGACAACCCCGGTACCACGCGCCACGAGGAACAGCACATATCTGATGATGAAGAGACAAC  300
Db   241  GAGCCCGACAACCCCGGTACCACGCGCCACGAGGAACAGCACATATCTGATGATGAAGAGACAAC  300
Qy   301  CCAAATAACAGAGAAAATGCAAAGATCTGCAATATAGCTTGAACTCTTTATCTAACAAAAC  360
Db   301  CCAAATAACAGAGAAAATGCAAAGATCTGCAATATAGCTTGAACTCTTTATCTAACAAAAC  360
Qy   361  GAGACGGTACTCTGAGGATGACTACCTCCAGACCATCACAAACATCAGAGATGCCCATG  420
Db   361  GAGACGGTACTCTGAGGATGACTACCTCCAGACCATCACAAACATCAGAGATGCCCATG  420
Qy   421  GAACCGGCAAGCAGAAGAATGACAATTTAGAGCAAAACTGGCTTCCTGTTGCGATGC  480
Db   421  GAACCGGCAAGCAGAAGAATGACAATTTAGAGCAAAACTGGCTTCCTGTTGCGATGC  480
Qy   481  CATTCAAGACTTCGTGGTTCCAGACAACACTCCAGTGGGACTAACATGAGCTACGA  540
Db   481  CATTCAAGACTTCGTGGTTCCAGACAACACTCCAGTGGGACTAACATGAGCTACGA  540
Qy   541  GGTGGAAAGCAAGAAACACATCCCCATTCGAGAGAACATTTTCCACATGTTTCCAGTGTC  600
Db   541  GGTGGAAAGCAAGAAACACATCCCCATTCGAGAGAACATTTTCCACATGTTTCCAGTGTC  600
Qy   601  GCAGCCTTTGTTGTTGGGACTATAACCCTATACCTGAAGTGCAGTGGTGGTAATGGGGAATTCT  660
Db   601  GCAGCCTTTGTTGTTGGGACTATAACCCTATACCTGAAGTGCAGTGGTGGTAATGGGGAATTCT  660
Qy   661  CAACAAGTCTCTCTGCGGAGCAGAAATGATAAATCGACTTCGTCTTCAGTGTAACCT  720
Db   661  CAACAAGTCTCTCTGCGGAGCAGAAATGATAAATCGACTTCGTCTTCAGTGTAACCT  720
Qy   721  CCCCCCAATCACAGGGAGCGCGCTAGTAAAGATGTTGGAAGCAAAACAAATCTTGTGACTGT  780
Db   721  CCCCCCAATCACAGGGAGCGCGCTAGTAAAGATGTTGGAAGCAAAACAAATCTTGTGACTGT  780
Qy   781  CAATCCCAGCAGCATTATAACCCTGAAGTACCAGAATTTGAAGGAGAAAAGCACAGTTTT  840
Db   781  CAATCCCAGCAGCATTATAACCCTGAAGTACCAGAATTTGAAGGAGAAAAGCACAGTTTT  840
Qy   841  GGAGGACACTCCACCTATGGAGATGCATTCCTGCCAGACATTTCGTCTGCTCTATCGGGC  900
Db   841  GGAGGACACTCCACCTATGGAGATGCATTCCTGCCAGACATTTCGTCTGCTCTATCGGGC  900
Qy   901  CAACACAGGCATCTCTTTAAAGTCTACCAAAACTCAAAGAGTCAAAAATGAGGCAAAA  960
Db   901  CAACACAGGCATCTCTTTAAAGTCTACCAAAACTCAAAGAGTCAAAAATGAGGCAAAA  960
Qy   961  GGTTCTCTTCTTCCATCCCAGTGCTACCTGAGACACACCTGCGCTCTTCTTCTGGAGAACTAAAGG  1020
Db   961  GGTTCTCTTCTTCCATCCCAGTGCTACCTGAGACACACCTGCGCTCTTCTTCTGGAGAACTAAAGG  1020
Qy  1021  GGTGACTGCATACCGCTTCGTGTCCAGGCTTGATGATTGCAAGTGTGCGTGTGAACTGTG  1080
Db  1021  GGTGACTGCATACCGCTTCGTGTCCAGGCTTGATGATTGCAAGTGTGCGTGTGAACTGTG  1080
Qy  1081  TGAAAACGTGAAGCTCACCGATTCTGGCCCTTTCTCTTAAGACTGGTTTCCACCAGATGGTTCCACCAGAT  1140
Db  1081  TGAAAACGTGAAGCTCACCGATTCTGGCCCTTTCTCTTAAGACTGGTTTCCACCAGAT  1140
Qy  1141  CAGTCACCACTACTACTAGATGATAACATGTTACCTAAGCATGTTCCTAAGATGAGAAGA  1200
Db  1141  CAGTCACCACTACTACTAGATGATAACATGTTACCTAAGCATGTTCCTAAGATGAGAAGA  1200
Qy  1201  ATACAGCCAAATGCTCCAGCTCCAGTCTCCAGCTCCATATGAGAGCAAATTCGCAATTCAGCAAATG  1260
Db  1201  ATACAGCCAAATGCTCCAGCTCCAGTCTCCAGCTCCATATGAGAGCAAATTCGCAATTCAGCAAATG  1260
Qy  1261  TGAAAACGGCTTAACGTTCTTAGAAGGAGAAGAGAATAATTTCAGGAGGTGGAGTGGATGTGTCA  1320
Db  1261  TGAAAACGGCTTAACGTTCTTAGAAGGAGAAGAGAATAATTTCAGGAGGTGGAGTGGATGTGTCA  1320
```

```
Qy  1321  CAGCATCTCCAAAAAGCCAATACGAGAGGGCACAGAGAAAGCATGAATTACAAAGGCGCT  1380
Db  1321  CAGCATCTCCAAAAAGCCAATACGAGAGGGCACAGAGAAAGCATGAATTACAAAGGCGCT  1380
Qy  1381  CTCCCACTTGTCTAGACCAAAGCCACCCACTTGCAGCCTCCACGAGTCA  1440
Db  1381  CTCCCACTTGTCTAGACCAAAGCCACCCACTTGCAGCCTCCACGAGTCA  1440
Qy  1441  CTCATTCTCACCTTCAACGTTCTTCTCTGAGAATAGAGACCAAAACATCAGACTTGGAT  1500
Db  1441  CTCATTCTCACCTTCAACGTTCTTCTCTGAGAATAGAGACCAAAACATCAGACTTGGAT  1500
Qy  1501  AAGTAAAAATGAGATAATTTTCAATCATCATAGAATTTGATTTGAGCCAGGGTCTCTCA  1560
Db  1501  AAGTAAAAATGAGATAATTTTCAATCATCATAGAATTTGATTTGAGCCAGGGTCTCTCA  1560
Qy  1561  GAATAGTCTTCCTTGTCTATCCGATAGCCAATTCCACCTTTATCAGAGTGGTAATGAA  1620
Db  1561  GAATAGTCTTCCTTGTCTATCCGATAGCCAATTCCACCTTTATCAGAGTGGTAATGAA  1620
Qy  1621  ACTGTGCAATTGTGCCAAAGACCCTTTCTGAAGAGAATGTCTGAATCATGCGCCGAGTTT  1680
Db  1621  ACTGTGCAATTGTGCCAAAGACCCTTTCTGAAGAGAATGTCTGAATCATGCGCCGAGTTT  1680
Qy  1681  TTACACAGAGCCTTCCCTTATAAATAAATCCTTCCCATTCCCTCCCTAGTAGAGTACA  1740
Db  1681  TTACACAGAGCCTTCCCTTATAAATAAATCCTTCCCATTCCCTCCCTAGTAGAGTACA  1740
Qy  1741  GAAACAAAATACCCTTCAGGAAGAAAAGTCTTTTACTTAGGCAATGCAATGTGCCTG  1800
Db  1741  GAAACAAAATACCCTTCAGGAAGAAAAGTCTTTTACTTAGGCAATGCAATGTGCCTG  1800
Qy  1801  CTTCTGATTCAGTTCGCTTGTGCATTAAGCTCGGTTGGTTTTGGGTTTGGGTTGGATTTGGGGC  1860
Db  1801  CTTCTGATTCAGTTCGCTTGTGCATTAAGCTCGGTTGGTTTTGGGTTTGGGTTGGATTTGGGGC  1860
Qy  1861  GTTTCTTCACTTCTTTTGTCTATAATTTCCTTACCTTGTTGTTATTCGAGCTTCC  1920
Db  1861  GTTTCTTCACTTCTTTTGTCTATAATTTCCTTACCTTGTTGTTATTCGAGCTTCC  1920
Qy  1921  TGCTTTGGGATTCGCAATTCTCCCCCACTGACAGGATCAACTCAATGACACATAAAGTA  1980
Db  1921  TGCTTTGGGATTCGCAATTCTCCTCCCACTGACAGGATCAACTCAATGACACATAAAGTA  1980
Qy  1981  GTTCAAACATCCATTCCATTGACATGTTTATCCATAAAGTTACTCATCGATTTTATT  2040
Db  1981  GTTCAAACATCCATTCCATTGACATGTTTATCCATAAAGTTACTCATCGATTTTATT  2040
Qy  2041  TAAAAATAGTGAACATCTACTTGACATAGACAGCCCGAGACCATCCTCCATTGGGAATATG  2100
Db  2041  TAAAAATAGTGAACATCTACTTGACATAGACAGCCCGAGACCATCCTCCATTGGGAATATG  2100
Qy  2101  AAGATATTGTCACTGGCAGAAAGCCAGTGTGTGCCATTAATTGATAAGATACCACCAAGC  2160
Db  2101  AAGATATTGTCACTGGCAGAAAGCCAGTGTGTGCCATTAATTGATAAGATACCACCAAGC  2160
Qy  2161  ATCATCATGCCAGTTATGAACACAGTGTCGTGAAAGGATCATAGACAGGGGGTGGTTAAATCT  2220
Db  2161  ATCATCATGCCAGTTATGAACACAGTGTCGTGAAAGGATCATAGACAGGGGGTGGTTAAATCT  2220
Qy  2221  GATCCCAGTAGAATAAACTTCAGGGCACATTAAGCCTAATTCACAATTAA  2280
Db  2221  GATCCCAGTAGAATAAACTTCAGGGCACATTAAGCCTAATTCACAATTAA  2280
Qy  2281  AACTAGTAAATGAACCAATTCTAGGCACATTAAGTGGATTCTGAGTAAAGAAGAGGGAA  2340
Db  2281  AACTAGTAAATGAACCAATTCTAGGCACATTAAGTGGATTCTGAGTAAAGAAGAGGGAA  2340
Qy  2341  CAGCAGGAGAAAGCTGTTCGCTTGGTTCTGATTCACCCAAATGAGCATGCTGGAAGGAGGT  2400
Db  2341  CAGCAGGAGAAAGCTGTTCGCTTGGTTCTGATTCACCCAAATGAGCATGCTGGAAGGAGGT  2400
```

```
US-10-501-930-4

Query Match            29.4%;        Score 930.2; DB 12; Length 1500;
Best Local Similarity  78.4%;        Pred. No. 1.5e-229;
Matches 1169; Conservative   0; Mismatches 308; Indels  14; Gaps   4;

Qy     2 GGAGCGGGCAGTCGGTCGGTCGCGCC.GGGGGCGTGGCGTTCGCCCCGGCAGCTTTCGGCGGGGAGGA  61
           |||||||||||||||||||||||||   |||||||||||||||||||||||| ||||| |||||||
Db    17 GGAGCCGCGGAGTCGGGTCGCGCC.GGGGGCGTGGCGTTCGCCCCGGCAGCTTTCGGCGGCGGG  76

Qy    62 CGCCCGTGGCTCAGGATGAGAT.CGGGGGCGCACGCGTGTTCGCGCCCTGTCTCGGCCAGCTCGATG 121
           ||||||||||||||||||||    ||||||||||||||||||||||| ||||||||||||||||
Db    77 CGCCTGTGGCTCAGGATGAGCGGCGGGGGCGCACTCGTGCTCGGCCCTGCTCGGCCAGCCTGCTG 136

Qy   122 CTGCTGCCTCCTGCGTATGC.CTGGTGCCAGCCGAGCGACGGCGCCTGCCCGCCTCCAGGCTG 181
           |||||||||||||||||||   |||||||||||||||||||||||||||||||||||||||||
Db   137 CTGCTGCTGCTGCGTGCGCGC.TGGTGCCGCGAGCAGCGGCGCCGGCGCGCCAGGGATT 196

Qy   182 TTGATGGAGGCAGAGAGA.CACCAGGTGTACCTGACTGGTACTGACTCCTGAGACACTCTGG 241
           |||||||||||||||||   |||||||||||||||||||||||||||||| |||||||||||
Db   197 CTGTGTGGAGGAAGCAGGGAGGCACCCACGGCACCCCGCAGGCGCTGAGGACGCTCCGG 256

Qy   242 AGCCGACAACCCGGTACCAC.CACCAGGAACAGCAGCACATCTGGATGAGAGACAACC 301
           |||||||||||||||||||   |||||||||||||||||||||||||||||||||||||
Db   257 AGCCGCGGACCGCGGTACCGC.CGCCACTAACAGCACATATCTGAATGAGAGTCGCTC 316

Qy   302 CAAATAACAGAGAAATGCAAAGA.TCTGCAATATAGCTTGAACTCTTTATCTAACAAACG 361
           ||||||||||||||||||||||   |||||||||||||||||||||||| |||||||||||
Db   317 CAACTGACCGAGAAATGCAAAA.TCGCAATATCGCATTCACAAACATCTCTCTAACAAACG 376

Qy   362 AGAGTACTCTGAGGATGAGT.CCTCAGACCATCACAAGCAGATGCCACGTGTCCATGG 421
           |||||||||||||||||||   |||||||||||||||||||||||||||| ||||||
Db   377 AAAGGGTATTCAGAGAACGACT.CCTCAGATTATCACAGATATCACAGAGTTGTCCATGG 436

Qy   422 AACCGGCAAGCAAGCAGAATATGA.CAATTTAGAGCAAAACTGGCTTCCTGTTCCGATGCC 481
           ||||||||||||||||||||||    |||||||||||||||||||| ||| ||| |||
Db   437 AAACGGCAAGCAAGCAGAATATGG.AAATTTTAGAGCCAAACTTGCTTCCTGCTGTGATGCT 496

Qy   482 ATTCAAGACTTCGTGGTTTCCA.GAACAACACTCCAGTGGGGACTAACATGAGCTACGAG 541
           ||||||||||||||||||||     |||||||||| |||||||| ||||||||||||||||
Db   497 GTTCAAAAACTTGTGTTGTTCTCA.GAATAACACTCCAGTTGGGACTAATATGAGTTACGAG 556

Qy   542 GTGGAAAGCAAGAACACATCCC.ATTCGAGAGAACACATTTCCACGTGTTCCAGTGTCG 601
           |||||||||||||||||||||   ||||||||||||||||||||| || |||||||||||
Db   557 GTGGAAAGCAAGCAAAAAGAAATCC.AATTAAGAAGACATATTTTCATATGTTTCCAGTGTCC 616

Qy   602 CAGCCTTTTGTGGGACTATCCCT.AATCAGTGTGCAGTGGTTGTAATGGGGAATTCTC 661
           |||||||||||||||||||||     ||||||||||| ||||||||||||||||||||||
Db   617 CAGCCTTTTGTGTGGACTACCCTT.ATAACAGTGTGCAGTGGTGTCGAAATGGGGAATTCTTG 676

Qy   662 AACAAGTCTCTCTGCGGAGCAGAG.AATTGATAAATCTGACTTCGTCTTCAGGTGTAACCTC 721
           |||||||||||||||||||||||   |||||||||||||||||||||||||||||||||||
Db   677 AATAAGTCTCTCTGTGGGAACTGA.ATAGATAAATCCGACTTCGTTTTTAGGTGTAACCTA 736

Qy   722 CCCCAATCACAGGGAGGGCCGCTAG.TAAAGATGTTGGAGCAAAACAAAATCTTGTGACTGTC 781
           ||||||||||||||||||||||||   |||||||||||||||||||||||||||||||||||
Db   737 CCCCAACCACAGGAGATGTTAGC.TAAAGATGTTGGCAGTAAAACAAAATCTTGTGACTATA 796

Qy   782 AATCCCAGCATGATTATAACCCTGA.GTACCAGAATTGAACGGAGAGAAAGCACAGTTTTG 841
           |||||||||||||||||||||||    ||||||||||||||||||||||||||||||||||
Db   797 AATCCAAGCATCATAACCTGAA.ATATGGGAACATATGGCAAAGGAAAAAAAGCCCATTTTTG 856

Qy   842 GAGGACATCTCCACCTATGGAGA.TGCATTCCTCCTGCCAGCATTTCCTATCGGGCC 901
           |||||||||||||||||||||    ||||||||||||||||||||||||||||||||
Db   857 GAGGACATTGCAACCTATGGAGA.GCATTTTTTTTCTGCCAGCATTTTCCTCAGGGCC 916

Qy   902 AACACAGGGTACATCTCTTTTAAAG.CTACCAAAACTCAAAGAGTCAAAAATGAGGCAAAAG 961
           |||||||||||||||||||||||   |||||||||||||||||||||||||||||||||||
Db   917 AACACGGGTACCTCTTTTCAAAG.ATACTACACGCTCGAAGAGCTCAAAGACAAGAACAAAG 976

Qy   962 GTTCTCTTCTTCCATCCCAGTTA.CTCGAGACACCTCGCTCTTTTTCTGGAGAACTAAAGGG 1021
           ||||||||||||||||||||||   |||||||||||| ||||||||||||||||||| ||||
Db   977 GTTCTATTTTTTCCATCCCAGTT.ACTACTACACGCTCGAAAGATCTCGGCCCCTTTTCTGGAGAACTAAAGGT 1036
```

```
Qy  2401 TGTGAGGCTACGCTAAAACCTCTGCGTAGGGAGGAGAGAGTACAGTGCATGAGTGTGGCGGCT 2460
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2401 TGTGAGGCTACGCTAAAACCTCTGCGTAGGGAGGAGAGAGTACAGTGCATGAGTGTGGCGGCT 2460

Qy  2461 TTTGTCCACACTCGTGAAGGGTGAGTAATTCAGAGCCAATCACATCACAAGGATGGACAC 2520
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2461 TTTGTCCACACTCGTGAAGGGTGAGTAATTCAGAGCCAATCACATCACAAGGATGGACAC 2520

Qy  2521 ACCTAACTCATCACTTCAGGGGAGGAGATGAATGCTTTCATGAGAAATTACACTCATAAGCT 2580
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2521 ACCTAACTCATCACTTCAGGGGAGGAGATGAATGCTTTCATGAGAAATTACACTCATAAGCT 2580

Qy  2581 AAGCACAGTTTGAGTAAAATTTGAGTAGATGTAAATATGAACATTTTATACCTCTTA 2640
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2581 AAGCACAGTTTGAGTAAAATTTGAGTAGATGTAAATATGAACATTTTATACCTCTTA 2640

Qy  2641 CTAATGTCCCACCGACACCTTTTAATGTAAGCACATTTATTTATTAAGTTACTTGACATT 2700
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2641 CTAATGTCCCACCGACACCTTTTAATGTAAGCACATTTATTTATTAAGTTACTTGACATT 2700

Qy  2701 AAATGCTTATGTCTGTATATTCGTTCATCCATTGATTTTCCCAAAAAGTAAGAGCATAG 2760
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2701 AAATGCTTATGTCTGTATATTCGTTCATCCATTGATTTTCCCAAAAAGTAAGAGCATAG 2760

Qy  2761 GAGATGAGGCCTACATGCCAAGAAAACTATAAATTTTACTCTTTAATTCTTTACTTGAGCC 2820
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2761 GAGATGAGGCCTACATGCCAAGAAAACTATAAATTTTACTCTTTAATTCTTTACTTGAGCC 2820

Qy  2821 AGCTTGTTGTTTATCAAGTGCTTTTTTGAGAGACAGCACCCTGTGAATTCTTCATTCTG 2880
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2821 AGCTTGTTGTTTATCAAGTGCTTTTTTGAGAGACAGCACCCTGTGAATTCTTCATTCTG 2880

Qy  2881 ATACAGTGTCACCTTGTATTTAACATTGTAATGTTGTTTACATCTCTTTCAT 2940
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2881 ATACAGTGTCACCTTGTATTTAACATTGTAATGTTGTTTACATCTCTTTCAT 2940

Qy  2941 TCTTTTATAGCAAATCAAACGTATTAGCTTCAGAAATTTATCAGAAGTTCATATATAAAT 3000
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2941 TCTTTTATAGCAAATCAAACGTATTAGCTTCAGAAATTTATCAGAAGTTCATATATAAAT 3000

Qy  3001 ATTTTGCAAAGGGTAAAAGGCTTTTTGTTTGTTAAATAAAATTATTATTTTCTTCTG 3060
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3001 ATTTTGCAAAGGGTAAAAGGCTTTTTGTTTGTTAAATAAAATTATTATTTTCTTCTG 3060

Qy  3061 ATGAATAGAGAGGCTCTTTATGCTGCTGCTAATGAACCTAATTGACTTTAAATTATCTCCT 3120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3061 ATGAATAGAGAGGCTCTTTATGCTGCTGCTAATGAACCTAATTGACTTTAAATTATCTCCT 3120

Qy  3121 AGCAACATTGGTCACGTTCAATCATGCTATTAGCAGCAAAAAAAAA 3166
         |||||||||||||||||||||||||||||||||||||||||||||||
Db  3121 AGCAACATTGGTCACGTTCAATCATGCTATTAGCAGCAAAAAAAAA 3166
```

```
Qy   1022 GTGACTGCATACCGCTTGTCCACAGGCTTGATGATTGCGCTGTGGAACTGTGT 1081
          ||||||||||||||||||||||||||  ||||||||||||| |||||
Db   1037 GTGACTGCATACCGCTTGTCCACCGGCTTGATGATCACAAGTGTTGCAGTGGAACTGTGT 1096

Qy   1082 GAAAACGTGAAGCTCTACGGATTCTGGCCTTTCTCTAAGACTATCGAAGACACACCCCACTC 1141
          ||||  |||||| | |||||| |||||||||||||||||  ||||||||||
Db   1097 AAAAATGTGAAGCTGTATGGATTCTGGCCCTTTCTAAAACTGAGAAGACATACCTGTC 1156

Qy   1142 AGTCACCACTACTATGATAACATGTTACCTTAAGCATGGTTTCCACCAGATGCTCAAAGAA 1201
          |||| ||| || ||| ||||||||||||||||||||  |||||||||||||||||||
Db   1157 AGCCATCACTATTATGACAACAGCTACCTAAACATGGTTTCCATCAGATGCCCAAAGAA 1216

Qy   1202 TACAGCCAAATGCTCCAGCTCCAATATGAGGAATCCTCAAACTCAAATGT 1261
          |||||||| ||| |||| | |||||||||||||||| ||| |||| |||||
Db   1217 TACAGCCAGATCCTCCAACTTCACATGAAAGGAATCCTCAAACTGCAATTAGCAAATGT 1276

Qy   1262 GAAACGGCTT----AACGTTTCTTAGAGGAGATAATTCACGA-GGTGGAGTGGATGT 1316
          |||||||||||      ||||||||||||||| ||||||||| ||||||||||||
Db   1277 GAAGTCGCCTAAACAAAGTATCTTAAAATGGGAATAATTTTAATATAATGCAGTAGGTGA 1336

Qy   1317 GTCACAGCATCTCCAAAAAGCCAATAGAAGAAGGCACAGAGAAAGCATGAATTACAAAGG 1376
          |||||||||||||||||||| |||||||||||||||  ||||||||||||||
Db   1337 TTAACAATGTCTCCAAACACC------AAAGGAGGTGGCTAAAGTATTTTGAGATG 1388

Qy   1377 CGCTCCCACTTGTCTAGACCAAAGCCACCCGCCCCCACTCACTTTGCAGCCTTCCACGA 1436
          ||||||||| |||  ||||||||||||||||| ||| ||||||||||  ||||
Db   1389 AGCCCCAAAATTTGGTTTGACCAAAGCTTCCCACTCATTTTGCAATGATGGCAAGTC-A 1447

Qy   1437 GTCACTCATTCTCACCTTCAAGGTTCTTTCTGTGAGAATAGAGACCAAAAC 1487
          |||||||| ||  ||||||||| ||||||||||||||||||  |||||||
Db   1448 TTCAATCCTTCTCATCTTCATTTTTTCTCCTTATAACATGGACACCACCATATC 1498
```

RESULT 3
US-10-301-480-562554/c
; Sequence 562554, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms
; TITLE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 562554
; LENGTH: 810
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-562554

Query Match   8.8%;   Score 278.2;   DB 12;   Length 810;
Best Local Similarity 65.5%;   Pred. No. 1.1e-60;
Matches   525;   Conservative   0;   Mismatches 248;   Indels 28;   Gaps   7;

```
Qy    982 GTACCTGAGACACCTGCTCTTTCTGGAGAACTAAAGGGGTGACTGCATACCGCTTGTC 1041
          ||||||||      |||||||||||||||||||||||| |||||||||||||||||
Db    810 GTACCTGAAAGATCTGGCCCTTTTCTGGAGAACTAAAGGTGTGACTGCATACCGCTTGTC 751

Qy   1042 CACAGAGGCTTGATGATTGCAAGTGTCGCTGCAGTGGACTGGAAACGTGAAGCTTACGG 1101
          |||| |||||||||||| | |||||| |||  |||||||| ||| |||||||||| |||
Db    750 CACCGGCTTTGATGATCACAAGTGTTGCAGTGGACTGTGTAAAAATGTGAAGCTGTATGG 691

Qy   1102 ATTCTGGCCTTTCTCTAAGAGATCTCTAAGAGATCGAAGACATCCAGTCAGTCACCACTACTATGATAA 1161
          |||||||||  |||||||||  ||  |||| ||  || ||||||| ||||| | ||||||
Db    690 ATTCTGGCCCTTTCTCTAAAACTGAGAGACATACCTGTCAGCCATCATTATTATGACAA 631
```

RESULT 4
US-10-301-480-1175963/c
; Sequence 1175963, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms
; TITLE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1175963
; LENGTH: 810
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-1175963

Query Match   8.8%;   Score 278.2;   DB 12;   Length 810;
Best Local Similarity 65.5%;   Pred. No. 1.1e-60;
Matches   525;   Conservative   0;   Mismatches 248;   Indels 28;   Gaps   7;

```
Qy    982 GTACCTGAGACACCTGCTCTTTCTGGAGAACTAAAGGGGGTGACTGACTACCGCTTGTC 1041
```

```
Qy   1162 CATGTTACCTTAAGCATGGTTTCACCAGATGCTCAAAGAATACAGCCAAATGCTCCAGCT 1221
          ||||| ||| ||| ||||||||||| ||| ||||||||||||||||||||||| || | |
Db    630 CAAGCTACCTAAACATGGTTTCGATCAGATGCCAAAGAATACAGCCAGATCCTCCAACT 571

Qy   1222 CCAATATGAGGAATCCTCAAAT GCAATTCAGCAAATGTGAAACGGCTT----AACGTT 1277
          |||  ||| |||||||||||||||||||| ||||||||||| | || |       |||||
Db    570 TCACATGAAAGGAATCCTCAAAT GCAATTTAGCAAATGTGAAGTCGCCTAAACAAAGTA 511

Qy   1278 TCTTAGAGGAGATAATTCACGA-GGTGGAGTGTGTCACAGCATCTCCAAAAAG 1336
          |||| || | ||||||||  ||   |||||  ||    ||||||||||||||
Db    510 TCTTAAAATGGGAATAATTTTAATATAATGCAGTAGGTGATTAACAATGTCTCCAAACAC 451

Qy   1337 CCAATAGAAGAAGGCACAGAGAAAGCATGAATTACAAAGGCGCTCCCACTTGTCTAGA 1396
          |      |||||||||||||||  ||||  ||||
Db    450 C-------AAAGGAGGTGGCTAAAGTATTTTGAGATGAGCCCCAAAATTTGGTTTGA 399

Qy   1397 CCAAAGCCACCCGCCCCCACTCACTTTGCAGCCTTCCACGAGTCACTC-ATTCTCACCTTC 1455
          ||||||      ||||||| | |||||| | |||  |||  ||   |  || || |||||
Db    398 CCAAAG------CTTCCCACTCATTTGCAATGATGGCAAGTCATTCAATCCTTCTCATC 344

Qy   1456 AAGGTTCTTTCTGTGAGAATAGCACCAAAACATCAGACTTGGATAAGTAAATGAGATA 1515
          ||  |||||| ||   ||| | ||||  ||||| |||||  ||| |||  ||   |||
Db    343 TTCATTTTTTCTCCTTATAACAGGACACCCATATCTGACTCATACAAATTTTATAATTT 284

Qy   1516 ATTTTCAAATCATCATAGAATTGATTTGAGCCAGGGTCTCTCAGAATGCTTCCTTGTT 1575
          ||  |||||| ||||  |||||  ||||| |  |||| || ||| |||| | ||| ||
Db    283 ATGAAAATTATTGGCATGGCCTTGGTGTGAGGTAGGGACTCAAAGAATATTTCTTTCTT 224

Qy   1576 CCTATCCATGATAGCCATTCCCACTTTATCAGAGTGGTAATGAAACTGTGCAATTGTGC 1635
          |||||   |||  ||| | |||  |||||  |   ||   ||  |  ||| |  |||  |
Db    223 ACTATTTATACCGCCATCTCCACCTCCACCAGACCGATAATAAAAGCATGCTTTCTTG- 165

Db    164 CAAAGACCCTTTCTGAAGAGAATCTGAAT------CATGCGCCGAGTTTTACACACA 1689
          |   ||  ||| || || ||||       |||||||      | ||| |||   |||
Db    164 --GAGATCTTCCTTATAAAATAAATCAGTCGCATTACCAGTATCCATTGGATTTCTTTCTGTG 107

Qy   1690 GCTCTTCCTTTATAAAATAAATCCTCCCATTCTCCCTCCTAGTAGTACAGAAGCAAAA 1749
          ||      | |  || ||| ||   ||||||||||||||  |  |||||||
Db    106 GCAGGTCTCTCATCAATAATATCCCCATTTTCTGTCTCAAAGACTGCATGCAAGAAA 47

Qy   1750 TACCCTTGATGATTCAGGAAG 1770
          |||||  |||| || | ||||
Db     46 ATTATATACTTAAGCAAAAAG 26
```

```
Query Match             8.0%;  Score 254;  DB 6;  Length 672;
Best Local Similarity   67.4%;  Pred. No. 1.8e-54;
Matches  437;  Conservative    1;  Mismatches  191;  Indels   19;  Gaps    5;

Qy    993 ACCTCGCTCTTTCTGGACAACCAAAGGGTGACTGCATACCGCTTGCCACAGGCTTGA 1052
              |||| ||||||||| | |||  ||||| | |||||||||||||  ||||| ||||||
Db     33 ATCTGGCCCTTTCTGTAGAACAAGGTGCGACTGCATACCGCTGTTCCACCGGCTTGA   92

Qy   1053 TGATTGCAAGTGTCGCTGTGGACTGTGTGAAAACGTGAAGCTCTCACGGATTCTGCCTT 1112
          |||| | ||||||| ||| ||||||||| |||| ||||||||  || ||||||| ||| |
Db     93 TGATCACAAGTGTTGCAGTGGACTGTGTAAAAATGTGAAGCTGTATGGATTCTGGCCCT  152

Qy   1113 TCTCTAAGACTATCGAAGACACCCCACTCAGTCACCACTACTATGATAACATGTTACCTA 1172
          ||||||| ||| | |||||||  |  ||||| |||||||  | ||  | ||| |  ||||||
Db    153 TCTCTAAAACTGTAGAAGACATCCTGTCAGCCATCACTATTATGACAACAAGCTACCTA  212

Qy   1173 AGCATGGTTTCCACCAGATGCCTAAAGAATACAGCCAAATGCTCCAGCTCCATATGAGAG 1232
          | ||||||||||| ||||||||  ||||||||||||| || ||||| ||||| |||| |||
Db    213 AACATGGTTTCCATCAGATGCCCAAAGAATACAGCCAGATCCTCCAACTTCACATGAAAG  272

Qy   1233 GAATCCTCAAACTGCAATTCAGCAAAATGTGAAACGGCTT----AACGTTTCTAGAAGGA 1288
          ||||||||||||||||||| ||| ||||||||||  |||| || ||||| ||| |||| ||
Db    273 GAATCCTCAAACTGCAATTTAGCAAATGTGAAGTCGCCTAAACAAAGTATCTTAAAATGG  332

Qy   1289 GAATAATTTCAGGA-GGTGGAGTGGATGTGTCACAGCATCTCCAAAAAGCCAATAGAAGA 1347
          ||||||||| |  |  | | |||   |||| | ||| ||||||||| | ||
Db    333 GAATAATTTTAATATAATGCAGTGGTGATTAACACATGTCTCCAAACACC--------AA  384

Qy   1348 AGGCACAGAGAAAGCATGAATTACAAAGGCGCTCTCCCACTTGTCTAGACCAAAGCCACC 1407
          ||| | | | ||||  | ||  ||    |  |||| ||  |  ||| |||| || |||
Db    385 AGGAGGTGGCTAAAGAGTATTYTSAGATGAGCCCCAAAATTTGGTTTGACCAAG-----  439

Qy   1408 CGCCCCCACTCACTTGCACCGCCTCCACAGTGCACTC-ATTCTCACGTTCTTTC 1466
          | |||| ||||| ||| |||| | |  | ||||||| | ||||||| |||||||||
Db    440 CTTCCCCACTCATTTTGCAATGATGGCAAGTCAGTGCATCAATCTTCATTTTTTC  499

Qy   1467 TCTGAGAATAGAGACCAAAACATAGACTTGGATAAGTAAATGAGATAATTTTTCAAAT 1526
          || | | |  | |||| | | || |||| ||||||| |||| | | |||  | ||| |
Db    500 TCCTATAACATGGACACCCATATCGACTCCATACAAATTTTATAATTTTATGAAAATTAT  559

Qy   1527 CATCATAGAATTTGATTTGAGCGGGTCTTCTCAGAATGCTTCCTTGTTCCATCCATGA 1586
          | |||| | ||||| || ||| || | |  |||||| | || ||| ||| |||||| |
Db    560 TGGCATGGCCCTTTGTCTGTGAGGTAGGGACTCAAAGAATATTTCTTTCTTACTATTTATAC  619

Qy   1587 TAGCCATTCCACCTTTATCAGATGGTAATGAAACTGTGCAATTGTG 1634
          |||||| | |||||| | ||| || | ||| ||| | |||| ||||
Db    620 CAGCCATCTCCACCTTCACCAGCCGATATTAAAAGCATGACTTTCTTG  667
```

```
Db    810 GTACCTGAAAGATCTGGCCCTTTCTGGAGAACTAAAGGTGTGACTGCATACCGCTTGTC  751
Qy   1042 CACAGGCTTGATGATTGCAAGTGTCGCTGTGGAACTGTGTGAAAACGTGAAGCTCTACGG 1101
Db    750 CACCGGCTTGATGATCACAAGTGTTGCAGTGGAACTGTGTAAAAATGTGAAGCTGTATGG  691
Qy   1102 ATTCTGGCCTTTCTCTAAGACTATCGAAGACACCCCACTCAGTCACCACTACTATGATAA 1161
Db    690 ATTCTGGCCCTTTCTCTAAAACTGTAGAAGACATCCTGTCAGCCATCACTATTATGACAA  631
Qy   1162 CATGTTACCTAAGCATGGTTTCCACCAGATGCCTAAAGAATACAGCCAAATGCTCCAGCT 1221
Db    630 CAAGCTACCTAAACATGGTTTCCATCAGATGCCCAAAGAATACAGCCAGATCCTCCAACT  571
Qy   1222 CCATATGAGAGGAATCCTCAAACTGCAATTCAGCAAAATGAAACGGCTT----AACGTT 1277
Db    570 TCACATGAAAGGAATCCTCAAACTGCAATTTAGCAAATGTGAAGTCGCCTAAACAAAGTA  511
Qy   1278 TCTAGAAGGAGAATAATTTCAGGA-GGTGGAGTGGATGTGTCACAGCATCTCCAAAAAG 1336
Db    510 TCTTAAAATGGGAATAATTTTAATATAATGCAGTGATTAACATGTCTCCAAACAC  451
Qy   1337 CCAATAGAAGGACAGAGAAAGCATGAATTACAAAGGCGCTCTCCCACTTGTCTAGA 1396
Db    450 C--------AAAGGAGGTGGCTAAAGAGTATTTTGAGATGAGCCCCAAAATTTGGTTTGA  399
Qy   1397 CCAAAGCCACCCGCCCCACTCACTTGCAGCGCTCCACGAGTCACTC-ATTCTCACCTTC 1455
Db    398 CCAAAG------CTTCCCCACTCATTTTGCAATGATGGCAAGTCAGTGCATCAATCTTCATC  344
Qy   1456 AACGTTCTTTCTCTGAGAATAGAGACCAAAACATCAGACTTGGATAAGTAAAATGAGATA 1515
Db    343 TTCATTTTCTCCTTATAACATGGACACCCATATCGACTCCATACAAATTTTTATATAATTT  284
Qy   1516 ATTTTTCAAATCATCATAGAATTTGATTTGAGCGCCAGGGTCTCTCAGAATGCTTCCTTGTT 1575
Db    283 ATGAAAATTATTGGCATGGCCCTTTGGTGTGAGGTGGGACTCAAAGAATATTTCTTTCTT  224
Qy   1576 CCTATCCATGATAGCCATTCCACCTTTATCAGAGTGGTAATGAAACTGTGCAATTGTGC 1635
Db    223 ACTATTTATACCAGCCATCTCCACCTTCACCAGCCGATAATAAAAGCATGCTTTCTTG-  165
Qy   1636 CAAAGACCCTTTCTGAAGAGAATGTCTGAAT------CATGCGCCGAGTTTTTACACACA 1689
Db    164 --GAGATCTCTTCTAAAGAGAATAGGATCCTGCATTAGTTACCAGTATCCATTTCTTCTGTG  107
Qy   1690 GCTCTTCCTTTATAAATAAATCCTTCCCATTCTCCCTCCTAGTAGAGTACAGAAAACAAA 1749
Db    106 GCAGGTCTCATCATAATATATCCCCATTTTTCTGTCTCAAAGACTGCATGCAAGAAA   47
Qy   1750 TACCCTTGATGATTCAGGAAG 1770
Db     46 ATTATATACTTTAAGCAAAAG   26
```

; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 44555
; LENGTH: 672
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-44555

Query Match   8.0%; Score 254; DB 7; Length 672;
Best Local Similarity 67.4%; Pred. No. 1.8e-54;
Matches   437; Conservative   1; Mismatches 191; Indels 19; Gaps 5;

```
Qy    993 ACCTCGCTCTTTTCTGGAGAACTAAAGGGGTGCATACCGGTCGTTGTCCACAGGCTTGA 1052
              ||| ||||||||||||||||||| |||||    ||||||||  |||||||  || ||||||
Db     33 ATCTGGCCCTTTCTGGTAGAACTAAAGGTGTGACTGCATACCGTGTTCCACCGGCCTTGA   92

Qy   1053 TGATTGCAAGTCGCGTGTGGAACTGTGTGAAAACGTGAAGCTCTACGGATTCTGGCCTT 1112
              |||| ||||| | | |||||||||||||| ||||| ||||||||| ||||||||||| |
Db     93 TGATCACAAGTGTTGCAGTGGAACTGTGTAAAAATGTGAAGCTGTATGGATTCTGGCCCT  152

Qy   1113 TCTCTAAGACTATCGAAGACACCCCACTCAGTCACCACTACTATGATAACATGTTACCTA 1172
              |||||||  || | |||||| ||| |||||||| |||||| |  ||||||||||||||||
Db    153 TCTCTAAAACTGTAGAGACATATCTGTCAGCCATCACTATATGACACAAGTACCTA      212

Qy   1173 AGCATGGTTTCCACCAGATGCCTAAAGAATACAGCCAAATGTCCAGCTCCATATGAGAG 1232
              | |||||||||| ||||||||| ||||||||||||||| |  ||||||||| ||| ||||
Db    213 AACATGGTTTCCATCAGATGCCCAAGAATACAGCCAGATCCTCCAACTTCACATGAAAG  272

Qy   1233 GAATCCTCAAACTGCAATTCAGCAAATGTGAAACGGCTT---AACGTTTCTTAGAAGGA 1288
              |||||||||||||||||| |||||||||||| ||| |||   ||| || |||| |||||
Db    273 GAATCCTCAAACTGCAATTTAGCAAATGTGAAGTGTCGCCTAAACAAAGTATCTTAAAATGG  332

Qy   1289 GAATAATTTCAGGA-GGTGGAGTGGATGGTGTCACAGCATCTCCAAAAAGCCAATAGAAGA 1347
              |||||||| | || || ||| | ||| | ||||||| ||||||||| |||------| |||
Db    333 GAATAATTTTAATATAATGCAGTAGGTGATTAACAATGTCTCCAAACACC-------AA  384

Qy   1348 AGGCACAGAGAAAGCATGAATTACAAAGCGCTCTCCCACTTGTCTAGACCAAAGCCACC 1407
              ||| || |||||||| |||||| |  |||||| ||||||| |  | |||||||-----
Db    385 AGGAGGTGGCTAAAGAGTATTYTGAGATGAGCCCAAAATTTGGTTTGACCAAAG-----  439

Qy   1408 CGCCCCACTCACTTTGCAGCCTCCAGAGTCACTC-ATTCTCACCTTCAACGTTCTTC 1466
              || |||||||| ||||||  | || |||| | || || |  ||  ||| | |||| ||
Db    440 CTTCCCCACTCATTTTGCACGTGGAGTGGCAAGTCATTCAATCTTCATCTCTCATTTTTTC  499

Qy   1467 TCTGAGAATAGAGACCAAAAACATCAGACTTGGATAAGTAAAATGAGATAATTTTCAAAT 1526
              || | |||| | || |||  || | |||||| || ||||  ||| |||  | |||||||
Db    500 TCCTTATAACATGGACAACCCATATCTGACTCATACAAATTTTTATAAATTTATGAAAATTAT  559

Qy   1527 CATCATAGAATTTGATTTGAGCCAGGGTCTCTCAGAATGCTTCCTTGTTCCTATCCATGA 1586
              || || ||  || | |||||| ||||| | ||  || ||| |  |||  | ||||||||
Db    560 TGGCATGGCCTTTGGTGTGAGGTAGGAGGGCTTTATTTGAAAGCATTGGAAAGTTCTGA  619

Qy   1587 TAGCCATTCCCCACCTTTATCAGAGTGGTAATGAGAAAACTGTGCAATTGTG 1634
              |||||| ||||||||| | |||||  ||  || |||||||| |  |||  ||
Db    620 CAGCCATCTCCACCTTCACCAGACCGGATATTAAAAGCATGGCTTTCCTTG  667
```

RESULT 7
US-10-085-117-66
; Sequence 66, Application US/10085117
; Publication No. US20030232334A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: 529520000121
; CURRENT APPLICATION NUMBER: US/10/085,117
; CURRENT FILING DATE: 2002-02-27
; PRIOR APPLICATION NUMBER: US 09/79,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 66
; LENGTH: 1071
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-085-117-66

Query Match   7.1%; Score 224.4; DB 7; Length 1071;
Best Local Similarity 55.3%; Pred. No. 1.2e-46;
Matches   457; Conservative   0; Mismatches 366; Indels   0; Gaps 1;

```
Qy    449 TTTAGAGCAAAACTGGCTTCCTCTGCGATGCCATTCAAGACTTCGTGGTTTCCAGAAC  508
              || ||| ||| ||| ||| ||||||||| |||||||  |||| | ||| | |||| |||
Db    232 TTCAGGAAACAAATGAAGACTCTGCGACCCTGCCATCTCTTTGCTATGACTAAAATG  291

Qy    509 AACACTCCAGTGGGACTAACATGAGCTACGAGGTGGAAAGCAAGAAACACATCCCCATT  568
              ||  |  |  |||||| | | |  |  | ||||| ||||| ||||| |||||||| |||
Db    292 AATTCCCCTATGGGAGAAGAGCATGTGGTATGACGGGGAGTTTTATACTTCATTCACCATT  351

Qy    569 CGAGAGAACATTTTCCACATGTTGTTGCAGTGTCGCAGCCTTTTGTGGACTATCCTATAAC  628
              | |  | ||| | |||| || |   |||   | |  |||||||  | ||| | ||||| |
Db    352 GACAATTCAACTTACTCTCTCTCCCACAGGCAACCCCAT---TCCAGTGCCATTGAAG  408

Qy    629 CAGTGTGCAGTGGTTGGTTAATGCGGAATTCTCAACAAGTCTCTCTGCCGGAGCAGAAATT  688
              ||||| | |||||| ||  ||| ||| |||||| |||| ||| ||| | || || | |||
Db    409 AAATGCGCGGTGGTGCGGGAAATGCGGGATTCTGGACAACATCTGGCTGTGGCCGTCAAATA  468

Qy    689 GATAAATCTGACTTCGTCTTCAGTGTAACCTCCCCAATCACAGGGAGCGCTAGTAAA  748
              |||  | |||| ||||| | ||  ||| ||| ||||| || || ||| | ||| |||
Db    469 GATGAAGCAAAATTTTGTCATGCCATGAAGTCACGCATTCCCAGTCAAGTGAATACACTAAG  528

Qy    749 GATGTTGGAAGCAAACAAATCTGTGACTGTCAATCCCAGCATTATAACCCTGAAGTAC  808
              ||||||||  | ||| | | | || |  | | || |||||| || | ||  | ||
Db    529 GATGTTGGATCCAAAAGTCAGTCAGTGACAGCAGCTAATCCCAGCATAATTCGGCAAAGGTTT  588

Qy    809 CAGAATTTGAAGAGGAGAGAAAGGACAGTTTTTGGAGGAGCATCTTCCACCTATGGAGATGCA  868
              ||||| ||  | ||  | ||||| |||||||| |||| || || |||||| |||  | ||
Db    589 CAGAACTTCTGTGGTCCAGAAGACATTTTGTGGACAACATGAAAATCTATAACCACAGT  648

Qy    869 TTCCTCCTCCTGCCAGCATTTCCTATCGGGCCAACAGCAGGCGATCTCTTTTAAAGTCTAC  928
              | | | ||| ||||  | ||||| |  |||| ||||| || | | ||| ||||| |||||
Db    649 TACATCTACTACACCTGCCTTTCGATGGGGAACAGAACAGAGAGCCATCTTGCAACCCAACTTTCTG  708

Qy    929 CAAACACTCAAAGAGTCAAAATGGAGGCAAAAGGTTCTCTTCCATCCCAGTCACCTG  988
              ||||||| |||||| || |  ||  | |  ||| | | |  || |||| ||||| ||
Db    709 TATACACTGTCAGATGTTGGTGCCAATCAAACAGTGCTGTTTGCAACCCAACTTTCTG  768

Qy    989 AGACACTTCGCTCTCTTTCTGACAGTACTAAAGGGGTGACTGACTGGTGTCCACAGGC 1048
              ||  | |||| | ||  |||| | | ||||||||||| |  |||  |||  ||||| |
Db    769 CGTAGCATTGGAAAGTTCTGGAAGTAGAGAATCCATGCCAAGCGCCGTCCACAGGA  828

Qy   1049 TTGATGATTGCAAGTGTCGCTGCGGAACTGTGTGAAAACGTGAAGCTCACGGATTCTCGG 1108
              |||| |||||||||| | |  || ||||||||||||||||||||| | ||||||||| ||
Db    829 CTTTTCTGGTGAGCGAGCTCCGGGTCTTCGTGTGAGGAGGTGGCCATCTATGGCTTCTCGG  888

Qy   1109 CCTTTCTCTAAGACTATCGAGAGACCCCACTCAGTCACCACTACTATGATAACATGTTA 1168
              || ||||| | || |||| | |||||||||| |||||||||||||||||| ||| | |||
Db    889 CCCTTCTCGTGAATATGCATGAGCAGCCCATCAGCACCACTACTATGACAACGTCTTA  948
```

Qy   869 TTCCTCCCTGCGCATTTCTATGGGCCAACACAGGCATCTCTTTAAGTCTAC 928
               |||||| ||||| ||| |||
Db   649 TACATCTACATGCGCCTTTCTATGAAGAACAGAGCCATCTTTGAGGGTTTAT 708

Qy   929 CAAACACTCAAAGAGTCAAAAATGAGGCAAAAGTTCTCTTCCTCCATCCCAGTACCTG 988
               | |||||
Db   709 TATACACTGTCAGATGTTGGTGCAATCAAACAGTGCTGTTTGCCAACCCAACTTTCTG 768

Qy   989 AGACACCTCGCTCTTTCTGGAGACTACTAAAGGGTGACTGCATACCGCTGTCCACAGGC 1048
Db   769 CGTAGCATTGGAAAGTTCTGGAAAGATGAGGAATCCATGCCAAGCGCCTGTCCACAGGA 828

Qy 1049 TTGATGATTGCAAGTGTCGCTGTTGAACTGTGTGAAAACGTGAAGCTCTACGGATTCTGG 1108
Db   829 CTTTTTCTGGTGAGCGCAGCTCTGGGTCTTCGTGAAGAGGTGGCCATCTATGGCTTCTGG 888

Qy 1109 CCTTTCCTAAGCTAATCTCGAAGACACCCCACTCAGTCACCACTACTATGATAACATGTTA 1168
Db   889 CCCTTCTCTTGTGAATATGCATGAGCAGCCCATCAGCACCACTACTATGACAACGTCTTA 948

Qy 1169 CCTAAGCATGGTTTCCACCAGATGCCTAAAGAATACAGCCAAATGCTCCAGCTCCATATG 1228
Db   949 CCCTTTCTGGCTTCCATGCCATGCCCGAGGAATTTCTCCAACTCTGGTATCTTCATAAA 1008

Qy 1229 AGAGGAATCCTCAAACTGCAATCAGCAAATGTGAAACGGCTTAAC 1274
Db 1009 ATCGGTGCACTGAGAATGCAGCTGGACCCCATGTGAAGATAGTACCTCAC 1054

RESULT 8
US-10-384-339C-136
; Sequence 136, Application US/10384339C
; Publication No. US20040175703A1
; GENERAL INFORMATION:
; APPLICANT: Kreutzer, Roland
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR INHIBITING EXPRESSION OF A TARGET GE
; FILE REFERENCE: 20200/2002
; CURRENT APPLICATION NUMBER: US/10/384,339C
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: PCT/EP02/00152
; PRIOR FILING DATE: 2002-01-09
; PRIOR APPLICATION NUMBER: DE 10100586.5
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: DE 10155280.7
; PRIOR FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: DE 10158411.3
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: DE 10160151.4
; PRIOR FILING DATE: 2001-12-07
; NUMBER OF SEQ ID NOS: 173
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 136
;   LENGTH: 1071
;   TYPE: DNA
;   ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
;   TITLE: GD3 synthase
;   PATENT DOCUMENT NUMBER: NM003034
US-10-384-339C-136

Query Match   7.1%; Score 224.4; DB 8; Length 1071;
Best Local Similarity 55.3%; Pred. No. 1.2e-46;
Matches 457; Conservative 0; Mismatches 366; Indels 3; Gaps 1;

Qy 449 TTTAGAGCAAAACTGGCTTCCTCGCGATGCCATTCAAGACTTCGTGGTTCCCAGAAC 508
              ||||| ||||||||| ||| || | ||||| |||||||| || |||||
Db 232 TTCAGGAAACAAATGAAGACTGCTGCGACCCTGCCCATCTCTTTGCTATGACTAAAATG 291

Qy 509 AACACTCCAGTGGGGACTAACATGAGCTACAGAGGTGGAAAGCAAGAAACACATCCCCATT 568
Db 292 AATTCCCCTATGGGGAAGAGCATGTGTATGACGCGGGAGTGTTTTATACTCATTCATTCACCATT 351

Qy 569 CGAGAGAACAGATTTTCCACATGTTTCCAGTGTCGCAGCCTTTTGTGTGGACTATCCTATAAC 628
Db 352 GACAATTCAACTTACTTCTCTCTTCCCCCAGGCAACCCCAT---TCCAGCTGCCATTGAAG 408

Qy 629 CAGTGTGCAGTGGTTGGTAATGGGGAATTCCAACAAGTCTCTCTGCGGAGCAGAAATT 688
Db 409 AAATGCGCGGTGTGTGGGAAAATGTGGGATTCTGGAAAGTGGGGCTGTGGCCGTCAAATA 468

Qy 689 GATAAATCTGACTTCGTCTTCAGGTGTAACCTCCCCCAATCACAGGGAGCGCTAGTAA 748
Db 469 GATGAGCAGCAAATTTGTCATGCGATGCAATCTCCCTCCTTTGTCAAGTGAATACACTAAG 528

Qy 749 GATGTTGGAAGCAAAGCCAAAAATCTTGTGACTGTCAATCCAGCATTATAAACCCTGAGTAC 808
Db 529 GATGTTGGATCAAAAGTCAGTTAGTGACAGCAGTAATTCGGCATAAATTCGGCAAAGTTT 588

Qy 809 CAGAATTTGAACGAGAGAGAAAGCACAGTTTTTGGAGGAACATCTCCAACTCCATGGAGATGCA 868
Db 589 CAGAACCTTCTGTGGTCCAGAAGACATTTGTGGGACAACATGTGAAAATCTATAACCACAGT 648

Db 714 TTCAGGAAACAAATGGAAGACTCTGGCGACCCTGCCCATCTCTTTGCTGATGACTAAAATG 773
        |||||||||||||||||
Qy 509 AACACTCCAGTGGGAGCTAACATCAGCTACGAGTGTGAAAGCAAGAAACACATCCCCATT 568

Db 774 AATTCCCTATGGGAAGAGCAAGTGGTATGACGGGGAGTGTTTTATATCATTCACCATT 833
Qy 569 CGAGAGAACATTTCCACATGTCCAGTGTCGCAGCCTTTGTGGACTATCCTATAAC 628

Db 834 GACAATTCAACTTACTCTCTCTCCACAGGCAACCCAT--TCCAGTGCCATTGAAG 890
Qy 629 CAGTGTGCAGTGGTTGGTAATGGGGAATTCTCAACAAGTCTCTCGCGGAGCAGAAATT 688

Db 891 AAATGCGGCGGTGTGGGAAATGCGGGATTCTGAAGAGAGAGTGGCTGTGGCCGTCAAATA 950
Qy 689 GATAAATCTGACTTCGTCTTCAGGTGTAACCTCCCCCAATCACAGGGAGCGCTAGTAAA 748

Db 951 GATGAAGCAAATTTGTCATGCCAATCTCCCTCCTTGTCAAGTGAATACACTAAG 1010
Qy 749 GATGTTGAAGCAAAACAAATCTGTGACTGTCAATCCCAGCATTATAACCCTGAAGTAC 808

Db 1011 GATGTTGGATCCAAAGTCAGTAGTGACAGCTAATCCCAGCATAATTCGGCAAAGGTTT 1070
Qy 809 CAGAATTTGAAGGAGAAGAAGCACAGTTTTTGGAGGACATCTCCACCTATGGAGATGCA 868

Db 1071 CAGAACCTTCTGTGGTCCAGAAACATTTGTGACACATGAAAAATCTATAACCACAGT 1130
Qy 869 TTCCTCCTCCTGCCAGCAATTTCTATCGGGCCAACACAGGCATCTCTTTAAAGTCTAC 928

Db 1131 TACATCTTACATGCCTGCCTTTTGTATGAAGACAGGAACAGAGCCATCTTTGAGGGTTTAT 1190
Qy 929 CAAACACTCAAAGAGTCAAAAATAGGCAAAAGGTTCTCTTCCATCCCAGGTACCTG 988

Db 1191 TATACACTGTCAGATGTTGGTGCCAATCAAACAGTGTTTGCCACCCCAACTTTCTG 1250
Qy 989 AGACACCTGCCTCTTTTCTGGAACTAAAGGGGTGACTGCATACCGGCTTGTCCACAGGC 1048

Db 1251 CGTAGCATTGGAAAGTTCTGGAAAAGTAGAGGAATCCATGCCAAGCGGCCTGTCCACAGGA 1310
Qy 1049 TTGATGATTGCAAGTGTCGCGTGGAACTGTGTGAAAAACTGAAGCTTCACGATTCTGG 1108

Db 1311 CTTTTCTCGGTGGTGAGCGCAGCTCGGGTCTCTGTGAAGAGGTGGCCATCTATGGCTTCTGG 1370
Qy 1109 CCTTTCTCAAGACTATCGAAGCGCCTAAGGATGGAATTTCTCCAACTCTCGGTATCTTCATAA 1168

Db 1371 CCCTTCTCTGTGAATATGCATGAGCAGCCCATCAGCCACCACTACTATGACAACGTCTTA 1430
Qy 1169 CCTAAGCATGGTTCCACCAGATGCCTAAAGAATACAGCCAAATGCTCCAGCTCCATATG 1228

Qy 1229 AGAGGAATCCTCAAACTGCAATCAGCAAATGTGAAACGGCTTAAC 1274
Db 1431 CCCTTTCTGGCTTCCATGCCATCAGCAAATGCTCCAGCTGCATAAA 1490

Db 1491 ATCGGTGCACTGAGAATGCAGCGGACCCCATGTGAGATACCTCAC 1536
Qy 1491 ATCGGTGCACTGAGAATGCAGCGGACCCCATGTGAGATACCTCAC 1536

Query Match   7.1%; Score 224.4; DB 7; Length 2117;
Best Local Similarity   55.3%; Pred. No. 1.8e-46;
Matches 457; Conservative 0; Mismatches 366; Indels 3; Gaps 1;

Qy 449 TTTAGACAAAACTGGCTTCCTGTTGCGATGCCATTCAAGACTTCGTTGGTTTCCCAGAAC 508

```
COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/10/430,325
      FILING DATE: 07-May-2003
      CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US/08/361,304A
      FILING DATE: 29-NOV-1994
      APPLICATION NUMBER: PCT/JP94/00495
      FILING DATE: 28-MAR-1994
      APPLICATION NUMBER: JP HEI-5-69988
      FILING DATE: 29-MAR-1993
ATTORNEY/AGENT INFORMATION:
      NAME: WILSON, MARY J.
      REGISTRATION NUMBER: 32,955
      REFERENCE/DOCKET NUMBER: 249-66
TELECOMMUNICATION INFORMATION:
      TELEPHONE: (703) 816-4000
      TELEFAX: (703) 816-4100
      TELEX: 200797 NIXN UR
INFORMATION FOR SEQ ID NO: 1:
   SEQUENCE CHARACTERISTICS:
      LENGTH: 2117
      TYPE: nucleic acid
      STRANDEDNESS: double
      TOPOLOGY: linear
   MOLECULE TYPE: cDNA to mRNA
   ORIGINAL SOURCE:
      ORGANISM: Homo sapiens
      STRAIN: WM266-4 cell
      CELL TYPE: melanoma
   SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-430-325-1

Query Match         7.1%;  Score 224.4;  DB 7;  Length 2117;
Best Local Similarity  55.3%;  Pred. No. 1.8e-46;
Matches  457;  Conservative  0;  Mismatches  366;  Indels  3;  Gaps  1;

Qy    449 TTTAGAGCAAAACTGGCTTCCTGTGCGATGCCATTCAGACTTCGTTCCCAGAC 508
              ||| ||||||||||||||||||||  | || ||||| || ||| |||||||
Db    714 TTCAGGAAACAAATGGAAGACTGCTGCGACCCTGCCCATCTCTTTGCTATGACTAAAATG 773

Qy    509 AACACTCCAGTGGGGACTAACATGAGCTACGAGGTGAAAGCAAGAAACACATCCCCATT 568
              || ||||||| ||||| |||||||||| ||||||||| |||||||||||||||||||
Db    774 AATTCCCCTATGGGAAGCATGTGGTATGACGGGAGTTTTTATACTCATTCACCATT 833

Qy    569 CGAGAGAACATTTTCCACATGTTTCCAGTGTCGCAGCCTTTTGTGGACTATCCCTATAAC 628
              |||||||||||  |||| |    |||   ||  | |   |  || |||||| |||
Db    834 GACAATTCAACTTACCTCTCTCTTCCCACAGGCAACCCCAT--TCCAGCTGCCATTGAAG 890

Qy    629 CAGTGTGCAGTGGTTGGTAATGGGAGGAATTCTCAACAGTCTCTCTGCGGAGCAGAAATT 688
              ||||||||||||||||| || |||  |||||||||| ||||| || ||||| ||  ||
Db    891 AAATGCCGCCGGTGGTGGGAAATGTGGGATTCTGAAGAAGAAGAGTGGCTGTGGCCGTCAAATA 950

Qy    689 GATAAAATCTGACTTCGTCTTCAGGTGTAACCTCCCCCAATCACAGGAGCGCTAGTAAA 748
              ||| ||||||||||  | |||||| |||||   |||||||| ||| |||| ||| ||
Db    951 GATGAAGCAAAATTTGTCATGCAATGCAATCTCCTCCTTGTCAAGTGAATACACTAAG 1010

Qy    749 GATGTTGAAGCAAAACAAATCTTGTGACTGTCAATCCCAGCATTATAACCCTGAAGTAC 808
              ||||||||||||||||   |||||  | || ||| ||||||| || ||||||||||
Db   1011 GATGTTGGATCCAAAGTCAGTTAGTGACAGCAGTAATCCCAGCATAAATTCGGCAAAGGTTT 1070

Qy    809 CAGAATTGAAGGAAGAAGAAAGCACAGTTTTTGTGGACAGCATCTCCACCTATGGAGATGCA 868
               ||||  | || || | | | |||||  |||| |||||| || ||  | |||||||
Db   1071 CAGAACCCTCTGTGGTCCAGAAAGACATTTGTTGGACAACATGAAAATCTATACCACCAGT 1130

Qy    869 TTCCTCCCTCGCCAGCATTTTCCTATCGGGCCAACAGGCATCTCTTTAAAGTCTAC 928
              ||| | ||||| |||    |||   | ||||   |  || ||  ||||| |||||
Db   1131 TACATCTCACATGCCACTTTTCTCATGAAGACAGAGGACGGACCATCTTGAGGGTTTAT 1190
```

```
; RESULT 12
; US-10-342-887-859
;     Sequence 859, Application US/10342887
;     Publication No. US20040058340A1
; GENERAL INFORMATION:
;     APPLICANT: Dai, Hongyue
;     APPLICANT: He, Yudong
;     APPLICANT: Linsley, Peter S.
;     APPLICANT: Mao, Mao
;     APPLICANT: Roberts, Christopher
;     APPLICANT: Van 't Veer, Laura Johanna
;     APPLICANT: Van de Vijver, Marc J
;     APPLICANT: Bernards, Rene
;     TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
;     FILE REFERENCE: 9301-188-999
;     CURRENT APPLICATION NUMBER: US/10,342,887
;     CURRENT FILING DATE: 2003-01-15
;     PRIOR APPLICATION NUMBER: 60/298,918
;     PRIOR APPLICATION FILING DATE: 2001-06-18
;     PRIOR APPLICATION NUMBER: 60/380,710
;     PRIOR APPLICATION FILING DATE: 2002-05-14
;     PRIOR APPLICATION NUMBER: 10/172,918
;     PRIOR APPLICATION FILING DATE: 2002-06-14
;     NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 859
;     LENGTH: 2117
;     TYPE: DNA
;     ORGANISM: Homo sapiens
US-10-342-887-859

Query Match         7.1%;  Score 224.4;  DB 8;  Length 2117;
Best Local Similarity  55.3%;  Pred. No. 1.8e-46;
Matches  457;  Conservative  0;  Mismatches  366;  Indels  3;  Gaps  1;

Qy    449 TTTAGAGCAAAACTGGCTTCCTGTGCGATGCCATTCAGACTTCGTTCCCAGAAC 508
              ||| ||||||||||||||||||||  | || ||||| || ||| |||||||
Db    714 TTCAGGAAACAAATGGAAGACTGCTGCGACCCTGCCCATCTCTTTGCTATGACTAAAATG 773

Qy    509 AACACTCCAGTGGGGACTAACATGAGCTACGAGGTGGAAAGCAAGAAACACATCCCCATT 568
              || ||||||| ||||| |||||||||| ||||||||| |||||||||||||||||||
Db    774 AATTCCCCTATGGGAAGCATGTGGTATGACGGGAGTTTTTATACTCATTCACCATT 833

Qy    569 CGAGAGAACATTTTCCACATGTTCCAGTGTCGCAGCCTTTTGTGGACTATCCCTATAAC 628
              |||||||||||  |||| |    |||   ||  | |   |  || |||||| |||
Db    834 GACAATTCAACTTACCTCTCTCTTCCCACAGGCAACCCCAT--TCCAGCTGCCATTGAAG 890

Qy    629 CAGTGTGCAGTGGTTGGTAATGGGAGGAATTCTCAACAGTCTCTCTGCGGAGCAGAAATT 688
```

Right column alignment

```
Qy  509 AACACTTCAGTGGGACTAACA GAGCTACGAGGTGGAAAGCAAGAAACAAGAAACACATCCCATT 568
        |  ||  ||||||||        |||  |||||          |||| ||  ||||
Db  247 AACTCCCCCATGGGGAAGAGCC GTGGTACGATGGGGAGCCTCTTATACTCGTTCACCATC 306

Qy  569 CGAGAGAACATTTCCACATGT GCCAGTGTCGCAGCCTTTGTGGACTATCGTTCCTATAAC 628
        |||  | |||||| | |||||   |||  ||          ----||
Db  307 GATAATTCACGTACTCCCTCT TCCCCAGGCAACCCCCT---TCCAGCTGCCATTGAAG 363

Qy  629 CAGTGTGCAGTGGTTGGTAATG GGGAATTCTCAACAGTCTCTCGCGGAGCAGAAATT 688
        ||  | |||||| |||||||   |||||||||
Db  364 AAATGTGCGGTGGTGGAAACGG TGGGATTCTGAAGATGAGTGGCTGTGCCCGTCAAATA 423

Qy  689 GATAAATCGACTCGTCTTCAG GTGTAACCTCCCCCAATCACAGGAGCGCTAGTAAA 748
        |||          |||||     | |||
Db  424 GATGAACCAAAATTTGTCATGC GGTCGAACCTGCTCCCTTGTCAAGTGAGTACACCAGA 483

Qy  749 GATGTTGGAAGCAAAACAAATC TGTGACTGTCAATCCCAGCATTATAACCTGAAGTAC 808
        |||||| |||||||          ||| |
Db  484 GATGTCGGTTCCAAAAACTCAGT GGTGACAGCTAACATAATTCGCCAGAGATTT 543

Qy  809 CAGAATTTGAAGGAGGAAGAAAGC ACAGTTTTTGGAGGACATCTCCACCTATGGAGATGCA 868
        |||||| |||  ||         |  ||||
Db  544 GAAAACCTGCTGTGGTCCAGAAA GAGTTTGTGGACAACATGAAGAATCTATAACCACAGT 603

Qy  869 TTCCTCCTCCTGCCAGCATTTTC CTATCGGGCCAACACAGGCATCTCTTTTAAAGTCTAC 928
        |||   | | |||||||  |  |   ||||
Db  604 TACATCTACACTGTGTGCCAATC AAACAGTGCGTGTTTGCCAACCCCAACTTTCTGTAT 663

Qy  929 CAAACACTCAAAGAGTGTGGTGCC AATCAAACAGTGCGTGTTTGCCAACCCCAACTTTCTT 988
        |||||||  ||  | |||  |||   ||
Db  664 TACACACTGAAAGATGTTGGAGCC AATCAAACAGTGCGTGTTTGCCAACCCCAACTTTCTT 723

Qy  989 AGACACCTCGCTCTTTCTGGAACT AAAGGGGTGACTACCGCGTTGTCCACAGGC 1048
        ||||||||||||||||||||||||  |||||||||||||||||||||||||||||
Db  724 CGGAACATTGGAAAGTTCTGAAGAG AGTAGGGGGATCCACGCCAAGCGCCTGTCCACAGGA 783

Qy 1049 TTGATGATTGCAAGTGTCGCGTGG GAACTGTGTGAAAACGTGAAGCTCTACGGATTCTGG 1108
        ||||||||||||||||||||||||  |||||||||||||||||||||||||||||||||||
Db  784 CTCTTTTTGGTGAGTGAGTGCAGCCT GGGCCTTCGTCATCTATGGGTGTCCATCTATGGCTTCTGG 843

Qy 1109 CCTTTTCTTAAGACTATCGAAGACA CAGCCAAAATGCTCCAGTCCATATG 1168
        |||||| |||||              |  |||
Db  844 CCCTTTTCTCGGTGAACATGCAAGGC GACCCTATCGTCACCACTACTATGACAAGTCTTG 903

Qy 1169 CCTAAGCATGGTTCCACCAGAT GCCTAAAAGAATACAGCCAAATGCTCCAGCTTGAATTCTTCACAA 1228
        ||   |  ||           |
Db  904 CCCTTCTCAGGCTACCATGCCAT GCCTGAGGAATTCCTTCAGCTTTGGTATCTTCACAA 963

Qy 1229 AGAGGAATCCTCAAACTGCAAT CAGCAAATGTGAAACGGCTTAAC 1274
        |||||||  |||||||    |||   ||||
Db 1229 AGAGGAATCCTCAAACTGCAAT CAGCAAATGTGAAACGGCTTAAC 1274

Qy      ATCGGGCGCGGCCTGAGGATGCAGC CGAACCCGGTGTGAGGAGCCGTCAC 1009
        ||||   |  |||||||
Db  964 ATCGGGCCGCGCCTGAGGATGCAGCC GGACCCGGTGTGAGGAGCCGTCAC 1009
```

RESULT 14
US-10-085-117-62
; Sequence 62, Application US/10085117
; Publication No. US20030232334A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: 529452000121
; CURRENT APPLICATION NUMBER: US/10/085,117
; CURRENT FILING DATE: 2002-02-27
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 62
; LENGTH: 1388
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-085-117-62

Left column alignment

```
Db  891 AAATGCGCGGTGGTGGGAAATGGTGGGATTCTGAAGAAGAGTGGCTGTGGGCCGTCAAATA 950
Qy  689 GATAAAATCGACTCGTCTTCAGGTGTAACCTCCCCCAATCACAGGAGCGCTAGTAAA 748
Db  951 GATGAAGCAAAATTTGTCATGCGATGCAATCTCCCTCCTTTGTCAAGTGAATACACTAAG 1010
Qy  749 GATGTTGGAAGCAAAACAAATCTTGTGACTGTCAATCCCAGCATTATAACCTGAAGTAC 808
Db 1011 GATGTTGGATCCAAAAGTCAGTTAGTTAGTGACAGCTAATTCGGCAAAGGTTT 1070
Qy  809 CAGAATTTGAAGGAGGAAGAAAGCACAGTTTTTGGAGGACATCTCCACCTATGGAGATGCA 868
Db 1071 CAGAACCTTCTGTGGTCCAGAAAGACATTTGTGGACACACATGAAAATCTATAACCACAGT 1130
Qy  869 TTCCTCCTCCTGCCAGCATTTTCCTATCGGGCCAACACAGGCATCTCTTTTAAAGTCTAC 928
Db 1131 TACATCTACACTGTGTGCCAATCAAACAGTGCGTGTTTGCCAACCCCAACTTTCTG 1190
Qy  929 CAAACACTCAAAGAGTGTGGTGCCAATCAAACAGTGCGTGTTTGCCAACCCCAACTTTCTG 988
Db 1191 TATACACTGTCAGATGTTGGTGCCAATCAAACAGTGCGTGTTTGCCAACCCCAACTTTCTG 1250
Qy  989 AGACACCTCGCTCTTTCTGGAACTAAAGGGGTGACTACCGCGTTGTCCACAGGC 1048
Db 1251 CGTAGCATTGGAAAGTTCTGAAAAGTAGAGGAATCCATGCCAAGCGCCTGTCCACAGGA 1310
Qy 1049 TTGATGATTGCAAGTGTCGCGTGGGAACGTGTGAAGCTGTACGGATTCTGG 1108
Db 1311 CTTTTTCTGGTGAGCGCAGCTCTGGGTCTCTGTGAAGAGGTGGCCATCTATGGCTTCTGG 1370
Qy 1109 CCTTTTCTTAAGACTATCGAAGACACACCCCATTCCAAGAATACAGCCAAATGCTCCAGTCCATATG 1168
Db 1371 CCCTTTCTCTGTGGTGAATATGCATGCCATGCCCGGAGGAATTTCCCAACTCTGTATCTTCATAAA 1430
Qy 1169 CCTAAGCATGGTTCCACCAGATGCCAAAATGCTCAGCTGCAAATGCTGAAACGGCTTAAC 1228
Db 1431 CCCTTTCTCGGTTCGGCTTCCATGCCATGTGAAACGGCTTAAC 1490
Qy 1229 AGAGGAATCCTCAAACTGCAATTCAGCAAATGTGAAACGGCTTAAC 1274
Db 1491 ATCGGGTGCACTGAGAAATGCAGCTGAGAATGCAGCCATGTGAAGATAACCTCAC 1536
```

RESULT 13
US-10-085-117-63
; Sequence 63, Application US/10085117
; Publication No. US20030232334A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: 529452000121
; CURRENT APPLICATION NUMBER: US/10/085,117
; CURRENT FILING DATE: 2002-02-27
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 63
; LENGTH: 1026
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-085-117-63

Query Match   6.6%; Score 210; DB 7; Length 1026;
Best Local Similarity 54.2%; Pred. No. 6.1e-43;
Matches 448; Conservative 0; Mismatches 375; Indels 3; Gaps 1;

```
Qy  449 TTTAGAGCAAAACTGGCTTCCTGTTGCGATGCATTCAAGACTTCGTTGGTTTCCCAGAAC 508
        ||| ||  ||||   ||  |   |||  |||  |   |||| ||||||| |||| |||||
Db  187 TTTCAGAGACAAATGGAAGACTTCGTTGACCCTGCCCATCTCTTTGCTATGACTAAGAATG 246
```

; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2140
; LENGTH: 1223
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-2140

Query Match   6.5%; Score 206.6; DB 16; Length 1223;
Best Local Similarity 53.7%; Pred. No. 5.2e-42;
Matches 441; Mismatches 373; Conservative   4; Indels   3; Gaps   1;

Qy   449 TTTAGAGCAAAACTGGCTTCCTGTTGCGATGCCATTCAAGACTTCGTGGTTTCCCAGAAC 508
          |||||| | |||||||||||||| |||||| ||||||||||||||||||||||||||||||
Db   280 TTCAGGAGACAAATGAAGACTGCTGCAACCCGCCATCTCTTTGCAATGACGAAGGGTG 339

Qy   509 AACACTCCAGTGGGGACTAACATGAGCTACGAGGTGGAAAGCAAGAAACACATCCCCATT 568
          |||||| ||||||||||||||||||
Db   340 AACTCCCCATGGGAAGAGCCGTGGTATGACGGGGAGTTTTATACTCGCTCACCATC 399

Qy   569 CGAGAGAACATTTTCCACATGTTCCAGTGTGCGAGCCTTTGTGGACTATCCCTATAAC 628
          | |||
Db   400 GACAMYTCCACGTACTCCCTYTCCCCCAGGCAACCCCCT---TCCAGCTGCCATTGAAG 456

Qy   629 CAGTGTGCAGTGGTTGGTAATGGGGAATTCTCAACAAGTCTCTCGCGGAGCAGAAATT 688
          || ||||||||| |||||||||||||||||
Db   457 AAATGTGCCGTGGTGGGGAAACGCGTGGGATTCTGAAGATGAGTGGCGTGGCGTCAAATA 516

Qy   689 GATAAATCTGACTTCGTCTTCAGGTGTAACCTCCCCCAATCACAGGGGCGCTAGTAAA 748
          |||||||||||||||| ||||||
Db   517 GATGAAGCAAATTTTGTCATGCGTGTAACCTTCCTGTCAAGTGAATACACCAGA 576

Qy   749 GATGTTGGAAGCAAAACAAATCTTGTGACTGTCAATCCCAGCATTATAACCTGAAGTAC 808
          ||||||||||||||||||||||||||||
Db   577 GATGTGGGTTCCAAGACTCAGTTGGTGCAGCTCAGTTGTGCAACATGAACATATAACCACAGT 636

Qy   809 CAGAATTTGAAGAGAGAAAGACACAGTTTTTGGAGGACATCTCCACCTATGGAGATGCA 868
          |||||| |||||||||||
Db   637 GAAAACCTGCTGTGGTCCAGAAAGAAGTTGTGGACAACATGAAGATCTATAACCACAGT 696

Qy   869 TTCCTCCTCCTGGCCAGCATTTTCCATCCGATCTCATCCCCAGGTACCTG 928
          ||||||||
Db   697 TTCCTCCCTGCGCAGCATTTTCTATGGGGCCAACAGCATCTCTTTTAAAGTCTAC 756

Qy   929 CAAACACTCAAAGAGTCAAAGACTGGGAATCAAACAGTGCTTTTTGCTAACCCAACTTTCTT 988
          |||||||||||||
Db   757 TACACCCTGAAAGACGTGGAGGCTGGAGCAATCAAACAGTCGTTTTTGCTAACCCAACTTTCTT 816

Qy   989 AGACACCTCGCTCTTTTCTGGAGAACTAAAGGGGTGACTGCATACCGCTGTCCACAGGC 1048
          |||||
Db   817 CGAAACATTGGAAAGTTCTGGAAGAGTTAGGGGATCCCAAGCGCTTGTCTCACAGGA 876

Qy   1049 TTGATGATTGCAAGTGCGCTGGGAACTGTGTGAAAACGTGAGCTCACGGATTCTGG 1108
          ||||||||||
Db   877 CTCTTTTTTGGGTGAGTGCAGCCTGGGAGGTGTGTCCATCTATGGCTTCTCGG 936

Qy   1109 CCTTTCCTTAAGACTATCGAAGACACCCCACTCACCACTACTATGATAACATGTTA 1168
          |||||||
Db   937 CCCTTCTCCGTGAACATGGAACATCGAAGCCGAACCTATCAGCCACCATTACTATTGACAAACGTCTTG 996

Qy   1169 CCTAAGCACATGGTTTCCACCCAGGATGCCAAAAGAATACAGCCAAATGCTCCAGCTCCATATG 1228
          |||||| | ||||||||| |||||
Db   997 CCCTTCTCAGGCTTCCATGCCATGCCYGGAGCAATTTCAACTCTGGTATCTTCATAAA 1056

Qy   1229 AGAGGAATCCTCAAACTGCAATTCAGCAAATGTGAAACGGC 1269
          ||| | | ||| ||||||| | : : |
Db   1057 ATGGGTGCCACTGGAGGATGCAACTGGACCCCATGTGAGRASCC 1097

Search completed: May 31, 2006, 20:45:31
Job time : 3710 secs

Query Match   6.6%; Score 210; DB 7; Length 1388;
Best Local Similarity 54.2%; Pred. No. 7.5e-43;
Matches 448; Conservative   0; Mismatches 375; Indels   3; Gaps   1;

Qy   449 TTTAGAGCAAAACTGGCTTCCTGTTGCGATGCCATTCAAGACTTCGTGGTTTCCCAGAAC 508
          |||||| | |||||||||||||| |||||| ||||||||||||||||||||||||||||||
Db   537 TTCAGGAGACAAATGAAGACTGCTGTGACCCTGCCATCTCTTTGCTATGACTAAGATG 596

Qy   509 AACACTCCAGTGGGGACTAACATGAGCTACGAGGTGGAAAGCAAGAAACACATCCCCATT 568
          |||||| ||||||||||||||||||
Db   597 AACTCCCCATGGGAAGAGCCGTGGTACGATGGGGAGCTCTTTATACTCGTTCACCATC 656

Qy   569 CGAGAGAACATTTTCCACATGTTCCAGTGTGCGAGCCTTTGTGGACTATCCCTATAAC 628
          | |||
Db   657 GATAATTCCACGTACTCCCTTCCCCCAGGCAACCCCCT---TCCAGCTGCCATTGAAG 713

Qy   629 CAGTGTGCAGTGGTTGGTAATGGGGAATTCTCAACAAGTCTCTCGCGGAGCAGAAATT 688
          || ||||||||| |||||||||||||||||
Db   714 AAATGTGCCGTGGTGGGGAAACGGTGGGATTCTGAAGATGAGTGGCCGTGTGCCCGTCAARATA 773

Qy   689 GATAAATCTGACTTCGTCTTCAGGTGTAACCTCCCCCAATCACAGGAGCGCTAGTAAA 748
          |||||||||||||||| ||||||
Db   774 GATGAACAAATTTTGTCATGCGGTGCAACCTGCCTCCTGTCAAGTGAGTACCACCAGA 833

Qy   749 GATGTTGGAAGCAAAACAAATCTTGTGACTGTCAATCCCAGCATTATAACCTGAAGTAC 808
          ||||||||||||||||||||||||||||
Db   834 GATGTCGGTTCCAAAACTCAGTTGGTGACAGCTAACCCCAGCATAATTCGCCAGAGATTT 893

Qy   809 CAGAATTTTGAAGAGAGAAAGACACAGTTTTTGGAGGACATCTCCACCTATGGAGATGCA 868
          |||||| |||||||||||
Db   894 GAAAACCTGCTGTGGTCCAGAAAGAAGTTGTGGACAACATGAAGATCTATAACCACAGT 953

Qy   869 TTCCTCCTCCTGGCCAGCATTTTCCATCCGATCTCATCCCCAGTACCTG 928
          ||||||||
Db   954 TACATTCACATGCCTGCCTTTTCTATGAAGACAGGCACAGAGCCGTCTCTCCGTGTGTAT 1013

Qy   929 CAAACACTCAAAGAGTCAAAGAATGAGGCAAAAAGTGCTTCTCTTCCATCCCAGGTACCTG 988
          ||||||||||||
Db   1014 TACACACTGAAAGATGTGTGGAGCCAATCAAACAACAGTGCTTTTTTGCTAACCCCAACTTTCTT 1073

Qy   989 AGACACCTCGCTCTTTTCTGGAGAACTAAAGGGGTGACTGCATACCGCTGTCCACAGGC 1048
          |||||
Db   1074 CGGAACATTGGAAAGTTCTGGAAGAGTTAGGGGATCCCAAGCGCTTGTCTACAGGA 1133

Qy   1049 TTGATGATTGCAAGTGCGCTGGGAACTGTGTGAAAACGTGAGCTCACGGATTCTGG 1108
          ||||||||||
Db   1134 CTCTTTTTTGGGTGAGTGCAGCCTGGGCCTGTGAGGAGGTGTCCATCTATGGCTTCTCGG 1193

Qy   1109 CCTTTCCTTAAGACTATCGAAGACACCCCACTCACCACTACTATGATAACATGTTA 1168
          |||||||
Db   1194 CCCTTCTCCGGTGAACATGGAACATCGAAGGGGACCCTATCAGTCACTAGTACTATGAACAACGTCTTG 1253

Qy   1169 CCTAAGCACATGGTTTCCACCCAGGATGCCAAAAGAATACAGCCAAATGCTCCAGCTCCATATG 1228
          |||||| | ||||||||| |||||
Db   1254 CCCTTCTCAGGCTTCCATGCCATGCCTGAGGAATTCCTTCAGCTTTGTATCTTCACAAA 1313

Qy   1229 AGAGGAATCCTCAAACTGCAATTCAGCAAATGTGAAACGGCTTAAC 1274
          ||| | | ||| ||||||| |
Db   1314 ATCGGCGCGCTGAGGATGCAGCTGGACCCGTGTGAGGAGCCGTCAC 1359

RESULT 15
US-11-136-527-2140
; Sequence 2140, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25

us-10-501-930-2.rnpbm

```
                    GenCore version 5.1.8
          Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on:   May 31, 2006, 19:15:45 ; Search time 50 Seconds
                          (without alignments)
                          7457.448 Million cell updates/sec

Title:           US-10-501-930-2
Perfect score:   3166
Sequence:        3166
    1 cggagcggcgagtcggtgcc.........gctattagcaaaaaaaaa 3166

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        246837 seqs, 58886990 residues

Total number of hits satisfying chosen parameters:     493674

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Published_Applications_NA_New:*
         1: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US09_NEW_PUB.seq:*
         2: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US06_NEW_PUB.seq:*
         3: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US07_NEW_PUB.seq:*
         4: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US08_NEW_PUB.seq:*
         5: /EMC_Celerra_SIDS3/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
         6: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10_NEW_PUB.seq:*
         7: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US11_NEW_PUB.seq:*
         8: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US60_NEW_PUB.seq:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	Query Match	Length	DB	ID	Description
c	1	42.8	1.4	1556	6	US-10-505-928-322	Sequence 322, App
c	2	39.8	1.3	541	6	US-10-488-619-949	Sequence 949, App
c	3	39.8	1.3	641	6	US-10-488-619-1577	Sequence 1577, Ap
c	4	39.2	1.2	785	6	US-10-953-349-3868	Sequence 3868, Ap
	5	39	1.2	1134	7	US-11-217-529-166160	Sequence 166160,
c	6	38.2	1.2	8391	7	US-11-217-529-166179	Sequence 166179,
c	7	38.2	1.2	70665	6	US-10-505-928-596	Sequence 596, App
c	8	38	1.2	3079	7	US-11-293-697-498	Sequence 498, App
c	9	36.6	1.2	1837	6	US-10-953-349-35182	Sequence 35182, A
c	10	36.2	1.1	1611	7	US-11-217-529-81129	Sequence 81129, A
c	11	36.2	1.1	2296	6	US-10-473-173-24	Sequence 24, Appl
c	12	36.2	1.1	37426	6	US-10-473-173-32	Sequence 32, Appl
	13	36	1.1	1241	6	US-10-953-349-22022	Sequence 22022, A
	14	36	1.1	4908	6	US-10-505-928-226	Sequence 226, App
	15	35.8	1.1	510	6	US-10-488-619-1510	Sequence 1510, Ap
c	16	35.6	1.1	1731	7	US-11-217-529-78208	Sequence 78208, A
c	17	35.4	1.1	785	6	US-10-946-650-15	Sequence 15, Appl
	18	35.4	1.1	2037	6	US-11-266-446-75	Sequence 75, Appl
c	19	35.4	1.1	3844	6	US-10-946-650-32	Sequence 32, Appl
c	20	35.2	1.1	2480	7	US-11-293-697-1939	Sequence 1939, Ap
	21	35	1.1	669	7	US-11-233-726-37	Sequence 37, Appl
c	22	35	1.1	1563	7	US-11-217-529-82562	Sequence 82562, A
c	23	35	1.1	2103	7	US-11-293-697-69	Sequence 69, Appl
c	24	34.8	1.1	300	6	US-10-488-619-2492	Sequence 2492, Ap
c	25	34.8	1.1	687	7	US-11-242-317-38	Sequence 38, Appl
c	26	34.8	1.1	2983	7	US-11-293-697-589	Sequence 589, App
	27	34.8	1.1	3288	6	US-10-196-749-477	Sequence 477, App
c	28	34.6	1.1	847	6	US-10-525-126-99	Sequence 99, Appl
	29	34.6	1.1	1887	7	US-11-217-529-4458	Sequence 4458, Ap
	30	34.4	1.1	1578	7	US-11-217-529-76837	Sequence 76837, A
c	31	34.4	1.1	5853	7	US-11-217-529-1030	Sequence 1030, Ap
c	32	34.2	1.1	1997	7	US-11-293-697-546	Sequence 546, App
c	33	34.2	1.1	2297	7	US-11-266-446-62	Sequence 62, Appl
c	34	34.2	1.1	2476	7	US-11-217-529-166168	Sequence 166168,
c	35	34.2	1.1	4941	6	US-10-713-648A-34	Sequence 34, Appl
c	36	34	1.1	600	7	US-11-217-529-166170	Sequence 166170,
	37	34	1.1	3460	6	US-10-293-697-1714	Sequence 1714, Ap
	38	33.8	1.1	958	6	US-10-505-928-315	Sequence 315, App
	39	33.8	1.1	1185	6	US-10-953-349-14579	Sequence 14579, A
	40	33.8	1.1	1599	6	US-10-953-349-27865	Sequence 27865, A
	41	33.8	1.1	1990	6	US-11-266-446-76	Sequence 76, Appl
c	42	33.8	1.1	2876	6	US-10-505-928-468	Sequence 468, Appl
	43	33.6	1.1	2884	7	US-11-293-697-700	Sequence 700, App
	44	33.4	1.1	600	7	US-11-217-529-166176	Sequence 166176,
	45	33.4	1.1	995	6	US-10-953-349-16178	Sequence 16178, A

ALIGNMENTS

```
RESULT 1
US-10-505-928-322
; Sequence 322, Application US/10505928
; Publication No. US20060088532A1
; GENERAL INFORMATION:
;  APPLICANT: Ludwig Institute for Cancer Research et al.
;  TITLE OF INVENTION: LYMPHATIC ENDOTHELIAL GENES
;  FILE REFERENCE: 28967/39178
;  CURRENT APPLICATION NUMBER: US/10,505,928
;  CURRENT FILING DATE: 2004-08-27
;  PRIOR APPLICATION NUMBER: US 60/363,019
;  PRIOR FILING DATE: 2002-03-07
;  NUMBER OF SEQ ID NOS: 866
;  SOFTWARE: PatentIn 3.2
;  SEQ ID NO 322
;  LENGTH: 1556
;  TYPE: DNA
;  ORGANISM: Homo sapiens
US-10-505-928-322

Query Match             1.4%;
Best Local Similarity   54.4%;  Score 42.8;  DB 6;  Length 1556;
Matches   86;  Conservative   0;  Mismatches  72;  Indels   0;  Gaps   0;    Pred. No. 0.069;

Qy    608 TTTGTGGACTATCCCTATAACCGTGTGCAGTGGTTGGTAATGGGGAATTCTCAACAAG 667
          |||| |||| |||||||||||| ||||||| |||||||||||||||| ||||| |||
Db    630 TTTGACACATACCCTGTAAAAATGTGTGGTGGTGGTAATGGGGAGTTTTGAAGAAT 689

Qy    668 TCTCTCTGCGGAGCAGAAATTGATAAATCTGACTTCGTCTTCAGGTGTAACCTCCCCCA 727
          ||||     ||||||| |||||| || ||      ||||   |            |||
Db    668 AAGACATTAGGAGAAAAATCGTCTCCTATGATGTAATATAAGAATGAATAATGGTCCT 749

Qy    728 ATCACAGGGAGGCCTAGTAAAGTTGTTGGAAGCAAAAC 765
          |||||||||||| |||||||| |||  ||||| |||||
Db    750 GTTTTAGGACATGAAGGAGAAGTTGGGAGAAGGACAAC 787

RESULT 2
US-10-488-619-949/c
; Sequence 949, Application US/10488619
; Publication No. US20060099578A1
; GENERAL INFORMATION:
;  APPLICANT: Greenlee, Winner and Sullivan, P.C.
;  TITLE OF INVENTION: Human Mitochondrial DNA Polymorphisms, Haplogroups, Associations
;  TITLE OF INVENTION: Physiological Conditions, And Genotyping Arrays
;  FILE REFERENCE: 98-01 WO
;  CURRENT APPLICATION NUMBER: US/10/488,619
;  CURRENT FILING DATE: 2004-03-01
```

```
;  NUMBER OF SEQ ID NOS: 3040
;  SOFTWARE: PatentIn version 3.1
;  SEQ ID NO 949
;  LENGTH: 541
;  TYPE: DNA
;  ORGANISM: Homo sapiens
US-10-488-619-949

Query Match             1.3%;  Score 39.8;  DB 6;  Length 541;
Best Local Similarity  48.1%;  Pred. No. 0.24;
Matches  113;  Conservative   0;  Mismatches  122;  Indels   0;  Gaps   0;

Qy   2842  TTTTTGAAGAGAGACAGCACCCTGTGAATTCTTCATTCTGATACAGTGTCACCTGTATTT 2901
                ||  |||||| ||  |||  |||| |||  |||||||||||||||||||||
Db    538  TGTATAAAAGGGACTGAGAAATTTATAAACTTTTTTTTCCTTACTGTCTTTTTTTCTAAAGTAA  479

Qy   2902  AACATTTGTAATGTGTTCAAGTTACATCTCTTTCATTCTTTTATAGCAAATCAAACG 2961
                |||| | | |||| ||  | |||||||||| || ||||  ||||||
Db    478  AAACAAAGAAATTATGTGCCAGATTTATGCATATATTTATGTTGCATAGAATAAAATT  419

Qy   2962  TATTAGCTTCAGAAATTTATCAGAAGTTCATATATAAATATTTTGCAAAGGGTAAAAGGC 3021
                |||||   |||   |  |||||   ||
Db    418  TTTAATCTTTAATTTTACATTTCCTAAATATAATATTTTAAGACGAAACATTTGTTCTATAGC  359

Qy   3022  TTTTTTGTTAAATAAAATAAAAATTATTATTTTCTTCTGATGAATGAGAGGCTCTT 3076
                ||||||  ||||| |  |||||||  ||||||
Db    358  TTTTCCCTTTTTAAGTAAGGAATTTATTTTTTTTTCTGAATTATTTTCTTCGT  304
```

```
RESULT 3
US-10-488-619-1577/c
;  Sequence 1577, Application US/10488619
;  Publication No. US20060099578A1
;  GENERAL INFORMATION:
;  APPLICANT: Greenlee, Winner and Sullivan, P.C.
;  TITLE OF INVENTION: Human Mitochondrial DNA Polymorphisms, Haplogroups, Associations
;  TITLE OF INVENTION: Physiological Conditions, And Genotyping Arrays
;  FILE REFERENCE: 98-01 WO
;  CURRENT APPLICATION NUMBER: US/10/488,619
;  CURRENT FILING DATE: 2004-03-01
;  NUMBER OF SEQ ID NOS: 3040
;  SOFTWARE: PatentIn version 3.1
;  SEQ ID NO 1577
;  LENGTH: 641
;  TYPE: DNA
;  ORGANISM: Mus musculus
US-10-488-619-1577

Query Match             1.3%;  Score 39.8;  DB 6;  Length 641;
Best Local Similarity  52.0%;  Pred. No. 0.26;
Matches   89;  Conservative   0;  Mismatches  82;  Indels   0;  Gaps   0;

Qy   1776  TCTTTTTTACTTAGCAATGGCCGCCTCGATTCAGTTCGCTTGTGACATTAAGCTGGG 1835
                | ||||||||||| ||| || |||||||||||||||||||||||||||||
Db    406  TTTTTTTTTTGTGTTTTTTTTTTTTCTTTTTTTTTTTTGTTTTTTTTTCCC  347

Qy   1836  TTGGGGTTTTGGTTGGATTTGGGGCGTTTCTTCACTCTTTTGTCTATATTTCCTTACC 1895
                |||||||||||  |||||| ||| ||||||||||||||||||||||
Db    346  TTGTTGTTTTGGTGTTTTTTTGTTTTCTGTGTTTTTTTTTTTTTTTTTTTTTTTTT  287

Qy   1896  TTTATCAGTTTGATTCGAGCTTCCTCGCTTGGGATTCTGCAATTCTCTCT 1946
                ||| ||||||||| ||||| |||||||
Db    286  TTTGTTTTGGTTTTTTTTGGTTTTTTTGGGTTTTTTTTCTTTGTTTTTCTTT  236
```

```
RESULT 4
US-10-953-349-3868
;  Sequence 3868, Application US/10953349
;  Publication No. US20060107345A1
;  GENERAL INFORMATION:
;  APPLICANT: ALEXANDROV, Nickolai et al.
;  TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
;  TITLE OF INVENTION: ENCONDED THERBY
```

```
;  FILE REFERENCE: 2750-1579PUS2
;  CURRENT APPLICATION NUMBER: US/10,953,349
;  CURRENT FILING DATE: 2004-09-30
;  NUMBER OF SEQ ID NOS: 40252
;  SOFTWARE: PatentIn version 3.3
;  SEQ ID NO 3868
;  LENGTH: 781
;  TYPE: DNA
;  ORGANISM: Arabidopsis thaliana
US-10-953-349-3868

Query Match             1.2%;  Score 39.2;  DB 6;  Length 781;
Best Local Similarity  53.2%;  Pred. No. 0.42;
Matches   83;  Conservative   0;  Mismatches  73;  Indels   0;  Gaps   0;

Qy   2902  AACATTTGTAATGTGTTCAAGTTACATCTCTTTCATTCTTTTATAGCAAATCAAACG 2961
                |||| | | |||| ||  | |||||||||| || ||||  |||||
Db    622  AATTTGTCTAATTCTCTCATTTTTCATTTCCTACCTTTTTTTCTTTTAAAAAAAT  681

Qy   2962  TATTAGCTTCAGAAATTTATCAGAAGTTCATATATAAATATTTTGCAAAGGGTAAAAGGC 3021
                |||||   |||   |  |||||
Db    682  TAATATGTACAGTCAGTTTTCTTATTTTCTTTCTTATACAATTTCAAACTTCAAAAACTC  741

Qy   3022  TTTTTTGTTAAATAAAATAAAAAT TATTATTTTCTT 3057
                |||||| ||||| |  |||||||
Db    742  CTTTTTTGGTAATCAAGAAATTCCACATTTTTTTT  777
```

```
RESULT 5
US-11-217-529-166160/c
;  Sequence 166160, Application US/11217529
;  Publication No. US20060099612A1
;  GENERAL INFORMATION:
;  APPLICANT: SUNTORY LIMITED
;  APPLICANT: NAKAO, YOSHIHIRO
;  APPLICANT: NAKAMURA, NORIHISA
;  APPLICANT: KODAMA, YUKIKO
;  APPLICANT: FUJIMURA, TOMOKO
;  APPLICANT: ASHIKARI, TOSHIHIKO
;  TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
;  FILE REFERENCE: S-38-285
;  CURRENT APPLICATION NUMBER: US/11,217,529
;  CURRENT FILING DATE: 2005-09-02
;  PRIOR APPLICATION NUMBER: US 10/912,182
;  PRIOR FILING DATE: 2004-09-02
;  NUMBER OF SEQ ID NOS: 197023
;  SOFTWARE: PatentIn version 3.3
;  SEQ ID NO 166160
;  LENGTH: 1134
;  TYPE: DNA
;  ORGANISM: Saccharomyces pastorianus
US-11-217-529-166160

Query Match             1.2%;  Score 39;  DB 7;  Length 1134;
Best Local Similarity  45.1%;  Pred. No. 0.59;
Matches  226;  Conservative   0;  Mismatches  270;  Indels   5;  Gaps   2;

Qy   2661  TTTAATGTAAGCACATTATTTACTTAAGTACTTGACATTAAATGCTTATGTCTGTATAT 2720
                ||||||||||  ||||||||||| || ||||||||||||| ||||||||
Db    801  TTTAATATTATTAATATTATTTATTATAGAAAATATTATGAAATCATT  742

Qy   2721  TCTGTTCATCCATCGATTTCCCAAAAAGTAAGAGCATTAGGAGATGAGGCCTACATGCCA 2780
                ||||   |||||||||||
Db    741  TAATTTAGGTATATATAATTTGTTATAATTCCATTAT-TAT  683

Qy   2781  AGAAAACTATAAATTTACTCTTAATTCTTACTTGAGCCAGCTTGTTGTTTATCAAGTG 2840
                ||| ||| || |||||||||| |  |||||||
Db    682  ATTTATCTATATCATTTAAACTAATATATTTACTAAAAATATCACTATTTATATAAATAT  623

Qy   2841  CTTTTTTGAAGAGACAGCAGCACCCTGAATTCTTCATTCTGATACAGTGTCACCCTGTATT 2900
                ||||| | || |  ||||||||||||||||||||||
Db    622  ATGATAATTTAATAGGTTCAATGTTACTCTTTTTTATTATATATAAATAATAACTAAAATATTAC  563
```

RESULT 6
US-11-217-529-166179
; Sequence 166179, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 166179
; LENGTH: 8391
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-166179

Query Match 1.2%;   Score 38.2;   DB 7;   Length 8391;
Best Local Similarity 47.7%;   Pred. No. 3;
Matches 112;   Conservative 0;   Mismatches 123;   Indels 0;   Gaps 0;

```
Qy 2846 TTGAAGAGACAGCACCCTGTGAATTCTTCATTCTGATACAGTGTCACCTTGTATTTAACA 2905
          |  | |||  || |    |||| |||| |||| |||||||| || ||  |||  ||||
Db 5238 TTAAAACAACACCTCCTTCTTATGTATTCCAATAAAATTAACCTAACTTTATTTATTTATAT 5297

Qy 2906 TTTGTAATGTGTTCAAGTTTACATCTCTTTCATTCTTTTATAGCAAATCAAACGTATT 2965
          |||  |||| || ||||||| |||| |||| |||||||| ||  ||| |||||
Db 5298 TATTTATAATCTTTATAATATTAATAATGATCTTTATTATATTCTTTTATTTAAAT 5357

Qy 2966 AGCTTCAGAAATTTATCAGAAGTTCATATATAAATATTTGCAAAGGGTAAAAGGCTTTT 3025
          ||  |  |||| ||||  |||||||||     |||| || |||| |||
Db 5358 ATATTATATATTGTTAATTATTATTATATTTGTTAATTATTATTT 5417

Qy 3026 TTGTTAAATAAATAAAATTATTATTTTCTTCTGATGAAATAGAGGCTCTTTTAT 3080
          | ||  |  |||  || |||| ||||  ||||
Db 5418 ATATTATTATTATTTAATTCTTTATATTTTATTAAGAATATATAAAGATGTTTTAT 5472
```

RESULT 7
US-10-505-928-596
; Sequence 596, Application US/10505928
; Publication No. US20060088532A1
; GENERAL INFORMATION:
; APPLICANT: Ludwig Institute for Cancer Research et al.
; TITLE OF INVENTION: LYMPHATIC ENDOTHELIAL GENES
; FILE REFERENCE: 28967/39178
; CURRENT APPLICATION NUMBER: US/10/505,928
; CURRENT FILING DATE: 2004-08-27
; PRIOR APPLICATION NUMBER: US 60/383,019
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 866
; SOFTWARE: PatentIn 3.2
; SEQ ID NO 596
; LENGTH: 70665
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-505-928-596

Query Match 1.2%;   Score 38.2;   DB 6;   Length 70665;
Best Local Similarity 52.1%;   Pred. No. 10;
Matches 85;   Conservative 0;   Mismatches 78;   Indels 0;   Gaps 0;

```
Qy 2868 ATTCTTCATTCTGATACAGTGTCACCTTGATTAACATTGTAATGTGTTCAAGTTTCAGTGT 2927
          |||||| |||| |||||||| || ||  |||  ||||  |||| || |||||||
Db 33814 ATATTTCAATTCAATTTGTGTCTACTTTATTGTCTTCAAATCAAACGTATTCAGAAAGTTTTCAGTGT 33873

Qy 2928 ACATCTCTTTCATTCTTTTATAGCAAATCAAACGTATTAGCTTCAGAAATTTATCAGAAG 2987
          ||| ||||||| |  | |  |||  ||| |||||
Db 33874 ACACCTTTTCATTTTCATTTTCATTTTTATCTTTTTGATGAT 33933

Qy 2988 TTCATATATAAATATTTGCAAAGGGTAAAAGGCTTTTTGTT 3030
          || ||||  |||  || | |||| |||
Db 33934 ATTTTAGATGAGATTTAGTAAGATGAATTGTTTCCCTATTTTTT 33976
```

RESULT 8
US-11-293-697-498/c
; Sequence 498, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/103,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 498
; LENGTH: 3079
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-498

Query Match 1.2%;   Score 38;   DB 7;   Length 3079;
Best Local Similarity 52.5%;   Pred. No. 1.9;
Matches 83;   Conservative 0;   Mismatches 75;   Indels 0;   Gaps 0;

```
Qy 2807 TTCTTACTTGAGCCAGCTTGTCTTTATCAAGTGCTTTTTGAAGAGACAGCACCCTGTG 2866
          || || ||  ||| |||| |||| |  ||  || ||||||
Db 2223 TTCATATTTTAGGAAGCTTGAAGCTGAAGCTAACTCAATATATATTTAAGACAATAAATCTCTTA 2164

Qy 2867 AATTCTTCATTCTGATACAGTGTCACCTGTATTTAACATTTGTTGTTTCAAGTT 2926
          |  ||||| ||||| |||||||||
Db 2163 TAATTTACAATAAATAAAGTGAAGTTGTATTATTCAAGACACAAGCAGAT 2104

Qy 2927 TACATCTCTTTCATTCTTTTATAGCAAATCAAACGTAT 2964
          ||  | || ||| |
Db 2103 TTTATCTGAAAAGACTTTTTTTTAAAATGAAGGCTTT 2066
```

RESULT 9
US-10-953-349-35182
; Sequence 35182, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:

```
Db    374 AAGTTTCTCTTGGTTATCATATATCTTCTGGGATTTCTTCTATGGAATTCATGATCTCC 315
                ||  |||  | ||||||||||||| |||||||| |||||  | ||| |||  ||||

Qy   3078 TATGCTGCTAATGAACCTAATTAGCTTAAATTA 3114
                ||||  ||| ||  ||   | |||||

Db    314 TCTATTTGTTCATAAAAATTTGAGAGCTTGTAATCA 278
```

; APPLICANT: ALEXANDROV, Nickolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: ENCODED THERBY
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 35182
;   LENGTH: 1837
;   TYPE: DNA
;   ORGANISM: Zea mays subsp. mays
US-10-953-349-35182

Query Match   1.2%; Score 36.6; DB 6; Length 1837;
Best Local Similarity   53.1%; Pred. No. 3.4;
Matches   78; Conservative   0; Mismatches   69; Indels   0; Gaps   0;

Qy   2 GGAGCGGCGAGTCGGTGCCGCCCCGGCGTGCGCTTCGCCCCGGCAGCTTTGGCGGCGAGGA 61
            |||| ||| || || | ||| | || | ||| | ||| ||| || ||||| |||
Db   156 GGATTGGCGGGGCTGAGGCAGCAGCAGCAGGCGAGCGTGGCCCCGTCGCGGCGGCAGCGCGGTG 215

Qy   62 CGCCCGTGGCTCAGGATGAGATCGGGGGGCACGCTGTTCGCCCTCATAGGCAGCCTGATG 121
            ||| |||||| | | | | ||| || ||| | || ||| ||| ||| || |
Db   216 GCCAGCGCGCTCGGCGGTGCGATCGCTGCGCGCGGCGGCGCCGGAGCCGCGTGG 275

Qy   122 CTGCTGCTCCTCCTGCGTATGCTCTGG 148
            |||||| | |||| | ||| || ||
Db   276 CGGCTTCCGCGGCCGGCGGCGGCGGCCGG 302

RESULT 10
US-11-217-529-81129/c
; Sequence 81129, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 81129
;   LENGTH: 1611
;   TYPE: DNA
;   ORGANISM: Saccharomyces pastorianus
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (16)..(16)
;   OTHER INFORMATION: a, c, g, t, unknown, or other
US-11-217-529-81129

Query Match   1.1%; Score 36.2; DB 7; Length 1611;
Best Local Similarity   47.9%; Pred. No. 4;
Matches 104; Conservative   0; Mismatches 113; Indels   0; Gaps   0;

Qy   2898 ATTTAACATTTGTAATGTTGTTTCAAGTTTACATTCTTTTATAGCAAATCA 2957
            ||| ||| || || || ||||| | || | |||| ||| | | ||
Db   494 ATTTTCAAATGCAAACTTTGTAAAGTCTAGCACTGTTTAAAGGTTCAATCTTCGCA 435

Qy   2958 AACGTATTAGCTTCAGAAATTTATCAGAAGTTCATATATAAAATATTTTGCAAAGGGTAAA 3017
            || ||||| ||||| | ||||| ||| |||| ||| |||| ||| || || ||| ||
Db   434 TCAGTTCCAGTGAAAGTGTTGTTAGGAAGAATATTTTTTTAGTCTTGCTAATGGATTC 375

Qy   3018 AGGCTTTTTGTTAAATAAATAAATTATTATTTTCTCTGATGAATAGAGAGGCTCTTT 3077
            |||| |||| |||||| || ||| ||||| |||| ||| ||| ||||| |||
```

```
Qy   3078 TATGCTGCTAATGAACCTAATTAGCTTAAATTA 3114
```

RESULT 11
US-10-473-173-24/c
; Sequence 24, Application US/1047317[?]
; Publication No. US20060088823A1
; GENERAL INFORMATION:
; APPLICANT: VAN ANDEL INSTITUTE
; TITLE OF INVENTION: Microarray Gene Expression Profiling in Clear Cell Renal Cell
; TITLE OF INVENTION: Carcinoma: Prognosis and Drug Target Identification
; FILE REFERENCE: 38345-170094
; CURRENT APPLICATION NUMBER: US/10/473,173
; CURRENT FILING DATE: 2003-09-29
; PRIOR APPLICATION NUMBER: US 60/279,411
; PRIOR FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 498
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 24
;   LENGTH: 2296
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-473-173-24

Query Match   1.1%; Score 36.2; DB 6; Length 2296;
Best Local Similarity   50.9%; Pred. No. 4.9;
Matches   86; Conservative   0; Mismatches   83; Indels   0; Gaps   0;

Qy   2 GGAGCGGCGAGTCGGTGCCGCCCCCGGCTGCGCTTGCGCTCCCCGGCAGCTTTGGCGGCGAGGA 61
            | || ||| ||| | ||| | || | | | | | || |||| ||| || ||||| |||
Db   1036 GGGAGGGAGGGGCGGGGTGCCGCGCGTCCCTTGTGCGGCGGGGCTGCGCCGG 977

Qy   62 CGCCCGTGGCTCAGGATGAGATCGGGGGGCACGCTGTTCGCCCTCATAGGCAGCCTGATG 121
            || ||||||| | | | | | ||| | || |||| || | | |||| || ||| |
Db   976 CTGCCGGGGCTCGGACCCCTCGGTGCTCGGGGCGCCGCGCGCCGCCTACGCCG 917

Qy   122 CTGCTGCTCCTCCTGCGTATGCTCTGG 170
            ||||| | | ||||| | || || ||| |
Db   916 CCGCTCCTGCGCCTCCGCTGCTCCGCGGCGGCGAGCTCGCCC 868

RESULT 12
US-10-473-173-32/c
; Sequence 32, Application US/1047317[?]
; Publication No. US20060088823A1
; GENERAL INFORMATION:
; APPLICANT: VAN ANDEL INSTITUTE
; TITLE OF INVENTION: Microarray Gene Expression Profiling in Clear Cell Renal Cell
; TITLE OF INVENTION: Carcinoma: Prognosis and Drug Target Identification
; FILE REFERENCE: 38345-170094
; CURRENT APPLICATION NUMBER: US/10/473,173
; CURRENT FILING DATE: 2003-09-29
; PRIOR APPLICATION NUMBER: US 60/279,411
; PRIOR FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 498
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 32
;   LENGTH: 37426
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-473-173-32

Query Match   1.1%; Score 36.2; DB 6; Length 37426;
Best Local Similarity   47.9%; Pred. No. 24;
Matches 104; Conservative   0; Mismatches 113; Indels   0; Gaps   0;

Qy   2888 GTCACCTTGTATTTAACATTTGTAAATGTTGTTTCAAGTTGTTTACATCTCTTTCATTCTTTA 2947
            |||| | | ||||||||||||| ||||||||||||||| ||||| | | | ||||
```

```
Db   18605 GGCACAGATGAATTAACATTACTGAGCATTATAGATTTTTTTTTTAATGGAGTGAATAT 18546
```

Matches   77; Conservative   0; Mismatches   50; Indels   1; Gaps   1;

```
Qy   2948 TAGCAAATCAAACGTATTAGCTTCAGAAATTTATCAGAAGTTCATATATAAATATTTTGC 3007
          ||| |||||||  | |||| | |||||||||||||||||||||||||||| || ||||||
Db   18545 AAGGATAATGCCAAATTTTAATTTTAAAAATGAATCACATATTCTTACTATATTTTTCTT 18486

Qy   3008 AAAGGGTAAAAGGCTTTTTTGTTAAATAAAATAAAAAATTATTATTTCCTCTGATGAATA 3067
          ||| ||| |||||||    ||  || ||| |||||||||||||| ||||| ||||
Db   18485 CAAAGTACTTCAAATTTTCTCTCTTTAAAAATTAATTCCTTCTTACAATGTTAAGGAAT 18426

Qy   3068 GAGGCTCTTTTATGCTGCTGCTAATGAACCTATTAG 3104
          |||||||| | || ||  |||||||||||||| ||
Db   18425 GAAGTTAGATCATTTATCTTTAAGAGTTGTAATCAG 18389
```

RESULT 13
US-10-953-349-22022
; Sequence 22022, Application US/10953349
; Publication No. US2006010734A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nickolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: ENCONDED THERBY
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 22022
; LENGTH: 1241
; TYPE: DNA
; ORGANISM: Glycine max
US-10-953-349-22022

Query Match   1.1%; Score 36; DB 6; Length 1241;
Best Local Similarity   51.9%; Pred. No. 3.9;
Matches   81; Conservative   0; Mismatches   75; Indels   0; Gaps   0;

```
Qy   2898 ATTTAACATTGTAAATGTTGTTCAAGTTTACATCTCTTCATTCTTTATAGCAAATCA 2957
          ||||  ||| || ||||  |||| |||| ||| |||| ||||||    |||| |||||
Db    351 ATTAATATTTATGATGAATTTTAGATTATAAATTATTTTTTATTTTTAGCCATTTC 410

Qy   2958 AACGTATTAGCTTCAGAAATTTATCAGAAGTTCATATATAAATATTTTGCAAAGGGTAA 3017
          |||| |||||||| |  ||||| ||||||| |||| |||||| ||||| ||||||||
Db    411 AATTTATTATTTATTTAAAACTGACATATTCAAGTAAAAAAAAAATTCATTCATGGGACAA 470

Qy   3018 AGGCTTTTTGTTAAATAAAATAAAATTTATTATTT 3053
          |||||||| ||||  ||||||||| |||| |||||
Db    471 AGAAAAATATATTTATAGAATAATTGTCCTTATT 506
```

RESULT 14
US-10-505-928-226
; Sequence 226, Application US/10505928
; Publication No. US2006008532A1
; GENERAL INFORMATION:
; APPLICANT: Ludwig Institute for Cancer Research et al.
; TITLE OF INVENTION: LYMPHATIC ENDOTHELIAL GENES
; FILE REFERENCE: 28967/39178
; CURRENT APPLICATION NUMBER: US/10/505,928
; CURRENT FILING DATE: 2004-08-27
; PRIOR APPLICATION NUMBER: US 60/363,019
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 866
; SOFTWARE: PatentIn 3.2
; SEQ ID NO 226
; LENGTH: 4908
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-505-928-226

Query Match   1.1%; Score 36; DB 6; Length 4908;
Best Local Similarity   60.2%; Pred. No. 8.5;
```

---

```
Qy   2881 ATACAGTGTCACCTGTATTTAA...ATTTGTAATGTTGTTCAAGTTACATCT-CTTTCA 2939
          ||||                     ||||||||||||||||||||| || |   | |
Db   3791 ATTGGGTTTCATTTGGGATTTGCAATGTAAATACGTATTTCTAGTTTTCATATAAAGTAG 3850

Qy   2940 TTCTTTATAGCAAATCAAACGCATAGCTTCAGAAATTATCAGAAGTTCATATATAAA 2999
          ||||||||      |||     |||||||||||||||||||||||||||||||||
Db   3851 TTCTTTTATAACAATGAAAAGTATTTTCTTGTATATTAAGTAAGTAATGAATATATAAG 3910

Qy   3000 TATTTTGC 3007
          ||||||||
Db   3911 AACTGTAC 3918
```

RESULT 15
US-10-488-619-1510
; Sequence 1510, Application US/10488619
; Publication No. US2006009978A1
; GENERAL INFORMATION:
; APPLICANT: Greenlee, Winner and Sullivan, P.C.
; TITLE OF INVENTION: Human Mitochondrial DNA Polymorphisms, Haplogroups, Associations
; TITLE OF INVENTION: Physiological Conditions, And Genotyping Arrays
; FILE REFERENCE: 98-01 WO
; CURRENT APPLICATION NUMBER: US/10/488,619
; CURRENT FILING DATE: 2004-03-01
; NUMBER OF SEQ ID NOS: 3040
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1510
; LENGTH: 510
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-488-619-1510

Query Match   1.1%; Score 35.8; DB 6; Length 510;
Best Local Similarity   52.3%; Pred. No. 2.6;
Matches   79; Conservative   0; Mismatches   72; Indels   0; Gaps   0;

```
Qy   2897 TATTTAACATTGTAAATGTTGTTCAAGTTTACATCTCTTCATTCTTTATAGCAAATC 2956
          ||||  ||| |||||||||| |||  ||||||||||  ||||  || |||||  |||
Db    259 TCTTAAACTGTTTAAATTTTCTAACATTTCCTATAGTGAAAGGCTATGAATTCAACC 318

Qy   2957 AAACGTATTAGCTTCAGAAATTTATCAGAAATTCATATATAAAATATTTGCAAAGGGTAA 3016
          ||| |||||||| || |||| |  |||| |||||||| |||  || |  ||||||||
Db    319 ATAGAAATTTCTCATTAAAACGTCAGTATTTCCTATTTTAGGATATTTGAAATTCTAA 378

Qy   3017 AAGGCTTTTTGTTAAATAAAATAAAATTTA 3047
          ||||  |||| ||||||||||| || |||
Db    379 AACGTTTTTTACTTAAATTAATTGTACTGA 409
```

Search completed: May 31, 2006, 19:18 56
Job time : 51 secs

```
               GenCore version 5.1.8
          Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on:    May 31, 2006, 11:50:53 ; Search time 568 Seconds
                                     (without alignments)
                                     10429.455 Million cell updates/sec

Title:         US-10-501-930-2
Perfect score: 3166
Sequence:      1 cggagcgcgagtcggtgcc........gctattagcaaaaaaaa 3166

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched:      1403666 seqs, 935554401 residues

Total number of hits satisfying chosen parameters:    2807332

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  Issued_Patents_NA:*
         1: /EMC_Celerra_SIDS3/ptodata/2/ina/1_COMB.seq:*
         2: /EMC_Celerra_SIDS3/ptodata/2/ina/5_COMB.seq:*
         3: /EMC_Celerra_SIDS3/ptodata/2/ina/6A_COMB.seq:*
         4: /EMC_Celerra_SIDS3/ptodata/2/ina/6B_COMB.seq:*
         5: /EMC_Celerra_SIDS3/ptodata/2/ina/7_COMB.seq:*
         6: /EMC_Celerra_SIDS3/ptodata/2/ina/H_COMB.seq:*
         7: /EMC_Celerra_SIDS3/ptodata/2/ina/PCTUS_COMB.seq:*
         8: /EMC_Celerra_SIDS3/ptodata/2/ina/PP_COMB.seq:*
         9: /EMC_Celerra_SIDS3/ptodata/2/ina/RE_COMB.seq:*
        10: /EMC_Celerra_SIDS3/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	224.4	7.1	1756	3	US-09-949-016-2130	Sequence 2130, Ap
2	224.4	7.1	2117	3	US-08-361-304A-1	Sequence 1, Appli
3	224.4	7.1	2117	3	US-09-949-016-436	Sequence 436, App
4	139.2	4.4	137394	3	US-09-949-016-13872	Sequence 13872, A
5	139.2	4.4	137743	3	US-09-949-016-12178	Sequence 12178, A
c 6	105.6	3.3	601	3	US-09-949-016-26413	Sequence 26413, A
c 7	105.6	3.3	601	3	US-09-949-016-72792	Sequence 72792, A
c 8	100.2	3.2	601	3	US-09-949-016-26414	Sequence 26414, A
c 9	100.2	3.2	601	3	US-09-949-016-72793	Sequence 72793, A
c 10	99.8	3.2	601	3	US-09-949-016-26415	Sequence 26415, A
c 11	99.8	3.2	601	3	US-09-949-016-72794	Sequence 72794, A
12	79.6	2.5	1048	2	US-08-626-994A-4	Sequence 4, Appli
13	79.6	2.5	1048	3	US-08-957-742-4	Sequence 4, Appli
14	79.6	2.5	1660	2	US-08-626-994A-2	Sequence 2, Appli
15	79.6	2.5	1660	3	US-08-957-742-2	Sequence 2, Appli
16	51.6	1.6	1292	2	US-08-503-133A-7	Sequence 7, Appli
17	51.6	1.6	1292	2	US-08-576-775A-7	Sequence 7, Appli
18	51.6	1.6	1292	3	US-08-972-498-7	Sequence 7, Appli
19	51.6	1.6	1292	3	US-08-899-545-7	Sequence 7, Appli
20	51.6	1.6	1895	3	US-09-949-016-4670	Sequence 4670, Ap
21	51.2	1.6	1682	2	US-08-503-133A-8	Sequence 8, Appli
22	51.2	1.6	1682	2	US-08-576-775A-8	Sequence 8, Appli
23	51.2	1.6	1682	2	US-08-972-498-8	Sequence 8, Appli
24	51.2	1.6	1682	3	US-08-899-545-8	Sequence 8, Appli
25	51	1.6	95890	3	US-09-949-016-16412	Sequence 16412, A
26	50.2	1.6	1141	3	US-09-806-708B-22	Sequence 22, Appl
27	50.2	1.6	20674	3	US-09-641-638-651	Sequence 651, App
28	50.2	1.6	20674	3	US-10-170-097-651	Sequence 651, App
c 29	49.8	1.6	7218	2	US-08-232-463-14	Sequence 14, Appl
c 30	49.2	1.6	7218	2	US-08-232-463-14	Sequence 14, Appl
c 31	49.2	1.5	50000	3	US-09-662-254B-23	Sequence 23, Appl
c 32	47.8	1.5	681	3	US-09-134-001C-741	Sequence 741, App
33	47.8	1.5	231129	3	US-09-949-016-16110	Sequence 16110, A
34	47.8	1.5	266293	3	US-09-949-016-11934	Sequence 11934, A
c 35	46.6	1.5	24922	3	US-09-949-016-13921	Sequence 13921, A
c 36	46	1.5	57605	3	US-09-949-016-13259	Sequence 13259, A
c 37	45.4	1.4	18651	3	US-09-949-002-592	Sequence 592, App
c 38	45.4	1.4	18682	3	US-09-949-002-786	Sequence 786, App
c 39	45.2	1.4	1141	3	US-09-806-708B-22	Sequence 22, Appl
c 40	45.2	1.4	640681	3	US-09-949-002-1280	Sequence 1, Appli
c 41	45	1.4	1039	3	US-09-902-540-1280	Sequence 1280, Ap
c 42	45	1.4	187169	3	US-09-949-016-12776	Sequence 12776, A
c 43	45	1.4	191569	3	US-09-949-016-15940	Sequence 15940, A
44	44.8	1.4	1156	3	US-09-949-016-3594	Sequence 3594, Ap
45	44.8	1.4	640681	3	US-09-790-988-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT 1
US-09-949-016-2130
; Sequence 2130, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE: 2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2130
;  LENGTH: 1756
;  TYPE: DNA
;  ORGANISM: Human
US-09-949-016-2130

Query Match             7.1%;  Score 224.4;  DB 3;  Length 1756;
Best Local Similarity   55.3%;  Pred. No. 2.9e-50;
Matches  457;  Conservative    0;  Mismatches  366;  Indels    3;  Gaps    1;

Qy       449 TTTAGAGCAAAACTGGCTTCTG|TGCGATGCCATTCAAGACTTCGTGGTTCCCAGAAC 508
                                    |      ||||     ||||||  |||||
Db       364 TTCAGGAAACAAATGGAAGACTCGTGCGACCCTGCCCATCTCTTGCTATGAGTAAAATG 423

Qy       509 AACACTTCCAGTGCGGGACTAACATCAGCTACGAGGTGAAAGCAAGAAACACATCCCCATT 568
                   ||||    |     ||||||                |||       ||| ||||||
Db       424 AATTCCCTATGGGGAAGAGCATGTGGTATGACGGGGAGTTTTATACTCATTCACCCATT 483

Qy       569 CGAGAGAACATTTCCACATGT|CCAGTGTCGCAGCCTTTTGTGGACTATCCCTATAAC 628
                         ||    |||        |       |              ||||
Db       484 GACAATTCAACTTACTTCTCTTCCACAGGCAACCCCAT---TCCAGCTGCCATTGAAG 540

Qy       629 CAGTGTGCAGTGGTTGGTAATGGGGAATTCAACAAGTCTTCTCTGCGGAGCAGAAATT 688
                          || |||      |||||||||          |    ||||
Db       541 AAATGCGCGGTGGTGGGAAATGCGGGGATTCGTGAAGAGAGTGGCTGTGGCCCTCAAATA 600

Qy       689 GATAAATCTGACTTCGTCTTCAGCGTGAACCTCCCCCAATCACAGGAGGCGTAGTAA 748
```

NAME: WILSON, MARY J.
REGISTRATION NUMBER: 32,955
REFERENCE/DOCKET NUMBER: 249-56
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
TELEX: 200797 NIXN UR
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2117
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
US-08-361-304A-1

Query Match   7.1%; Score 224.4; DB 3; Length 2117;
Best Local Similarity 55.3%; Prd. No. 3.2e-50;
Matches 457; Conservative 0; Mismatches 366; Indels 3; Gaps 1;

Qy   449 TTTAGAGCAAAACTGGCTTCCTGTGGCGATGGCCATTCAAGACTTCGTGGTTTCCCAGAAC 508
                     |||||| | ||| ||| |||||| | | ||||||| |||| ||||||
Db   714 TTCAGGAAACAAATGGAAGACTCCTGCGACCCTCCATCTTCTTGCTATGACTAAAAATG 773

Qy   509 AACACTCCAGTGGGGACTAACATAGCTACGAGGGTGGAAAGCAAGAAACACATCCCCATT 568
                       |||| ||| |||||||| | ||| ||||| |||||| | ||||||||||
Db   774 AATTCCCCTATGGGAAGACATGGTATGACGGGGAGTTTTTATATCATTCACCATT 833

Qy   569 CGAGAGAACATTTTCCACATGTTTCAGTGTCGCAGCCTTTTGTGGACTATCCCTATAAC 628
                       |||||| | ||||| |||||| ||| | ||| | ||| |||||
Db   834 GACAATTCAACTTACTCTCTGTCCCACAGGCAACCCCAT---TCCAGCTGCCATTGAAG 890

Qy   629 CAGTGTGCAGTGGTTGGTAATGGGGGAATTCTCAACAGTCTCTCGCCGGAGCAGCAGAATT 688
                      |||| || || |||||| ||||||| ||||| |||||| || || || ||||
Db   891 AAATGCGCGGTGGTGTGGGAAATGCGGGATTCTGAAGAAGAGTGGCTGTGCCGTCAAATA 950

Qy   689 GATAAATCTGACTTCGTCTTCAGCGTGTAACCTCCCCCAATCACAGGAGCGCTAGTAAA 748
                      |||||||| | || ||||| | ||||||| ||| || |||| | | |||
Db   951 GATGAAGCAAATTTGTCATGCGATGCAATCTCCCTCCTTTGTCAAGTGAATACACTAAG 1010

Qy   749 GATGTTGGAAGCAAAACAAATCTGTGACTGTCAATCCCAGCATTATAACCCTGAAGTAC 808
                      ||||||| | ||| |||||| ||||||||||| |||| |||| | || |||||||
Db   1011 GATGTTGGATCCAAAAGTCAGTCAGTGACAGCTAATCCCAGCATAATTCGGCAAAGGTTT 1070

Qy   809 CAGAATTTGAAGAGGAAGAAGCAGTTTTGGAACTGTGCTGTGTTCCACCTATGGAGATGCA 868
                      |||||||||| ||||||||| || | | |||| || || | ||||||||||||
Db   1071 CAGAACCTTCTGGTCCAGAAACACATTTGTGACAACATGAAAATCTATAACCACAGT 1130

Qy   869 TTCCTCCCTGCCAGCATTTTGTATCGGGCCAACACAGGACATCTCTTTTAAAGTCTAC 928
                      | |||||| ||| ||||| ||||| |||| ||||||| || | |||||| |||
Db   1131 TACATCTACATGCCTGCCTTTTCTATGAAGAACAGGAACAGAGCCATCTTTGAGGTTTAT 1190

Qy   929 CAAACACTCAAAGAGTCAAAAATGAGGCAAAAGGTTCTCTTCTCATCCCAGTACCTG 988
                      || || |||| |||| | |||| || ||||| || | | |||| || | |||
Db   1191 TATACACTGCAGATGTTGGTGCAATCAAACAGTGCTGTTTGCCACCCCAACTTTCTG 1250

Qy   989 AGACACCTCGCTCTTTTCTGACTGCTCATACCGGTTGTCATACCGGTTGTCCACAGGC 1048
                      || ||| ||| | || ||||| |||| ||| |||| |||| || || ||||||
Db   1251 CGTAGCATTGGAAAGTTCTGAAGTAGAGGAATCCATGCCAAGCGCCTGTCCACAGGA 1310

Qy   1049 TTGATGATTGCAAGTGTCGCTGTGGAACTGTGTGAAAACGTGAGCTCTACGGATTCTGG 1108
                      || || ||| || | |||||| ||| ||| || |||| | | ||| ||||||||||
Db   1311 CTTTTCTGGTGAGCCGAGCTCGGGTTCTGTGAAGAGGTGGCCATCTATGGGCTTCTGG 1370

Qy   1109 CCTTTCCTAAGACTATCGAAGACACCCCACTCACTACTATGATAACATGTTA 1168
                      || ||| | |||| ||| |||| |||||||| ||||| ||| |||| |||
Db   1371 CCCTTCCTCGTGAATATGACATGAGCCGCCTCATCAGCCACTACTATGAGTCTTA 1430

Qy   1169 CCTAAGCATGGTTTCCACCAGATGCCTAAAGAATACAGCCAAATGCTCCAGCTCCATATG 1228
                      || || || | |||||| || || | ||||||||| ||||| || || || || |||
Db   1431 CCCTTTTCTGGCTTCCATGCCATGCCCCAGGAATTTCTCCAACTCTGGTATCTTCATAAA 1490

Qy   1229 AGAGGAATCCTCAAACTGCAATTCAGCAAATGTGAAACGGCTTAAC 1274

RESULT 3
US-09-949-016-436
; Sequence 436, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 436
; LENGTH: 2117
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-436

Query Match   7.1%;   Score 224.4; DB 3;   Length 2117;
Best Local Similarity   55.3%;   Pred. No. 3.2e-50;
Matches 457;   Conservative   0;   Mismatches 366;   Indels   3;   Gaps   1;

Qy   449 TTTAGAGCAAAACTGGCTTCCTGTTGCGATGCCATTCAGAGACTTCGTGTTTCCCAGAAC 508
          ||||||| | |||||||||| ||||||| ||||||||| || | ||||| |||||
Db   714 TTCAGGAAACAAATGGAAGACTGCTGCGACCCCTGCCCATCTCTTTGCTATGACTAAAATG 773

Qy   509 AACACTCCAGTGGGGACTAAACATGAGCTACGAGGTGAAAGCAAGAAACACATCCCCATT 568
          ||||||| | |||| || ||| || ||| | | |||| ||||||||
Db   774 AATTCCCCTATGGGAGCAGCATGTGTATGACCGGGGAGTTTTTATACTCATTCACCATT 833

Qy   569 CGAGAGAACATTTTCCACATGTTTCGGACTGTCCGCAGCCTTTGTGGACTATCCCTATAAC 628
          |||| ||||||| | |||||| |||| || | | |||| || |||||
Db   834 GACAATTCAACTTACTCTCTCTTCCCACAGCCAACCCAT---TCCAGCTGCCATTGGAG 890

Qy   629 CAGTGTGCAGTGGTTGGGTAATGGGGAATTCTCAACAAGTCTCTGCGGAGCAGAAATT 688
          ||||| | || |||| ||| | || ||||| | ||| | || | || |||| ||||
Db   891 AAATGCGCGGTGCGTGGTGTGGGAAATGTGGGATTCTGAAGAAGAAGAAGTGGCTGTGGCCGTGAAATA 950

Qy   689 GATAAAATCTGACTTCGTCTTCAGGTGTAACCTCCCCCAATCACAGGAGACGGCTAGTAAA 748
          |||||||||||| || | |||| ||||| |||||||| ||||| || | || |||||
Db   951 GATGAAGCAAAATTTGTCATGCGAAGCAATCTCCCTCCTTGTCAAGTGAAATACACTAAG 1010

Qy   749 GATGTTGAAGCAAAAACAAAATCTTGTGACTGTCAATCCAGCATTATAACCCTGAAGTAC 808
          |||||||| | ||||||| | |||| || |||| ||||| | ||||| |||||
Db   1011 GATGTTGATCCAAAGTCAGTTAGTGACAGCTAATCCCAGCATAATTCGGCAAAGGTTT 1070

Qy   809 CAGAATTTGAAGGAGAAGAAAGCACACAGTTTTGGAGGACATCTCCACCTATGGAGATGCA 868
          |||| | ||| | | |||| || | |||||| |||| | | |||| |||
Db   1071 CAGAACCTTCTGTGGTCAGAAAGACATTTGTTGGACAACACATGAAAATCTATAACCACACGT 1130

Qy   869 TTCCTCCTCTGCCAGCATTTTCCTATCGGGCCAACAACAGGCATCTCTTTTAAAGTCTAC 928
          |||| | ||| | | | || ||||| |||| || || ||| |||| ||||| || |
Db   1131 TACATCTACACATGCCTGCCTTTTCTATGAAGACAGGAACAGAGCCCATCTTTGAGGGTTTAT 1190

Qy   929 CAAACACTCAAAGAGTCAAAAATGAGGCAAAAGGTTCTCTTCTTCCATCCCAGGTACCTG 988
          ||| || ||||| | || |||| | || |||||||| ||| |||||||| | ||
Db   1191 TATACACTGTCAGATGTTGGTGCCAATCAAACAGTGCTGTTTTGCCAACCCCAACTTTCTG 1250

Qy   989 AGACACCTCGCTCTTTTCTGGAGAACTAAAGGGGGTGACTAAGGGGTCGCATACCCGCTTGTCCCACAGGC 1048
          |||| | ||| | | | ||||| | |||||| | || |||||| || || ||||||| |||||||
Db   1251 CGTAGCATTGGAAAGTTCTGAAAAGTAGAGAGATAGAGAAATCCATGCCAAGGCGCCTGTCCCACAGGA 1310

RESULT 4
US-09-949-016-13872
; Sequence 13872, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13872
; LENGTH: 137394
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(137394)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13872

Query Match   4.4%;   Score 139.2; DB 3;   Length 137394;
Best Local Similarity   55.9%;   Pred. No. 4.4e-26;
Matches 264;   Conservative   0;   Mismatches 208;   Indels   0;   Gaps   0;

Qy   803 AAGTACCAGAATTGAAGGAGAAGAAAGCACAGTTTTGGAGGACATCTCCACCTATGGA 862
          |||| | | |||| | || ||||||| |||| ||||| ||| |||||||| |||||
Db 134353 AGGTTTCAGAACCTTCTGTGGTCAGAGAAAGACATTGTGACAACATGAAAATCTATAAC 134412

Qy   863 GATGCATTCCTCCTCCTGCCAGCATTTTCCTATCGGGCCAACACAGCAGGCCATCATCTTTAAA 922
          |||||||| | || | | |||| || | ||||| |||| |||| |||| || || |||||
Db 134413 CACAGTTACATCTACACATGCCTGCCTTTTCTATGAAGACAGGAACAGAGCCATCTTTGAGG 134472

Qy   923 GTCTACCAAACACTCAAAGAGTCAAAAATGAGGCAATCAAACAGTGTTTGCCAACCCCAAC 982
          ||| | ||| || |||| | || ||| |||| || ||||| | || | || ||||
Db 134473 GTTTATTATACACTGTCAGTATCAGAATGTGGTGCCAATGAAACAGTGCTGTTTTGCCAACCCCAAC 134532

Qy   983 TACCTGAGACACCTCGCTCTTTTCTGGAGAACTAAAGGGGTGACTACCGCTTGTCC 1042
          ||| | || | || | | | | |||||| | |||| ||||| | ||||||||
Db 134533 TTTCTGCGTAGCATTGGAAAGTTCTGAAAAGTAGAGAGATGGGAGCCGCGCTGTCC 134592

Qy   1043 ACAGGCTTGAATGATTGCAAGTGGCTGTGGAACGTGAAAACGTGAAGCTCTACGGA 1102
          |||| | ||| | | || ||| | | |||| | |||| || | || |||||||
Db 134593 ACAGGACTTTTTCTCGGTGAGCGGCGTCTGGGTGCTCTGTGAAGAGGTGGCCATCTATGGC 134652

Qy   1103 TTCTGGCCTTTCTAAGACTATCGTGTGAATATCATGAGCGACTATCACCACTACTATGATAAC 1162
          ||||||||||| | ||| || | | ||||| | ||| | | | | |||||| ||||||
Db 134653 TTCTGGCCCTTCTCTGTGAATATCATGAGCAGCCATCACCACTACTATGACAAC 134712

Qy   1163 ATGTTACCTAAGCATGGTTTCCACCAGATGCCTAAAGAATACAGCCAAATGCTCCAGTC 1222
              |||||||| |||||||| |||||| |||| || || || ||
Db   134713 GTCTTACCCCTTTCTGGCTTCCATGCCATGCCCGAGGAATTTCTCCAACTCTGGTATCTT 134772

Qy   1223 CATATGAGAGGAATCCTCAAACTGCAATTCAGCAAATGGAAACGGCTTAAC 1274
              |||||| | ||||||| || |||
Db   134773 CATAAATCGGTGCACTGAGAATGCAGCTGGACCCATGTGAGACCTCAC 134824

RESULT 5
US-09-949-016-12178
; Sequence 12178, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;   APPLICANT: VENTER, J. Craig et al.
;   TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;   TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;   FILE REFERENCE: CL001307
;   CURRENT APPLICATION NUMBER: US/09/949,016
;   CURRENT FILING DATE: 2000-04-14
;   PRIOR APPLICATION NUMBER: 60/241,755
;   PRIOR FILING DATE: 2000-10-20
;   PRIOR APPLICATION NUMBER: 60/237,768
;   PRIOR FILING DATE: 2000-10-03
;   PRIOR APPLICATION NUMBER: 60/231,498
;   PRIOR FILING DATE: 2000-09-08
;   NUMBER OF SEQ ID NOS: 207012
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12178
;   LENGTH: 137743
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(137743)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12178

Query Match   4.4%;   Score 139.2;   DB 3;   Length 137743;
Best Local Similarity   55.9%;   Pred. No. 4.4e-26;
Matches 264; Conservative   0; Mismatches 208; Indels   0; Gaps   0;

Qy   803 AAGTACCAGAATTGAAGGAGAAGCACAGTTCCCACCTATGGA 862
             || |||| |||||||||||||||||||||||||||
Db   134704 AGTTTTCAGAACCTTCTGTGGTCAGAAAGACATTTGTGGACAACATGAAAATCTATAAC 134763

Qy   863 GATGCATTCCTCCTCCTGCCAGCATTTCCTATCGGGCCAACACACAGGCATCTCTTTTAAA 922
             |||||||||||||| |||||||||||||||||| |
Db   134764 CACAGTTACATCTACATCTACATGCCTGCTTTTCTATGAAGACAGGAACAGGAGCCATCTTTGAGG 134823

Qy   923 GTCTACCAAACACTCAAAGAGTCAAAATGAGGCAAAAGGTTCTTCTTCCATCCCAGG 982
             |||||||||||||||||| |||
Db   134824 GTTTATTATACACTGTCAGATGTGGTGCCAATCAAACAGTGCTGTTTGCCAACCCCAAC 134883

Qy   983 TACCTGAGACACCTGCTCTTTTCTGAGAACTAAAGGGGTGACTGCATACCGCTTGTCC 1042
             ||||||||||||||||||||||||||||||||||
Db   134884 TTTCTGGTAGCATTGGAAAGTTCTGGAAAAGTAGAGGAATCCATGCCAAGCGCCTGTCC 134943

Qy   1043 ACAGGCTTGATGATTGCAAGTGCCGCTGTGGAACTGTGTGAAAACGTGAGCTCTACGGA 1102
              ||||||||||||||||||||||||||| |||||||||||||
Db   135004 TTCTGGCCTTTCTCTAAGACTATCGAAGACAACCCACCACTACTATGATAAC 135063

Qy   1163 ATGTTACCTAAGCATGGTTTCCACCAGATGCCTAAAGAATACAGCCAAATGCTCCAGTC 1222
              ||||||| |||||||| |||||| |||| || || || ||
Db   135064 GTCTTACCCCTTTCTGGCTTCCATGCCATGCCCGAGGAATTTCTCCAACTCTGGTATCTT 135123

Qy   1223 CATATGAGAGGAATCCTCAAACTGCAATTCAGCAAATGGAAACGGCTTAAC 1274
              |||||| | ||||||| || |||

Db   135124 CATAAATCGGTGCACTGAGAATGCAGCTGGACCCATGTGAGACCTCAC 135175

RESULT 6
US-09-949-016-26413/c
; Sequence 26413, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;   APPLICANT: VENTER, J. Craig et al.
;   TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;   TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;   FILE REFERENCE: CL001307
;   CURRENT APPLICATION NUMBER: US/09/949,016
;   CURRENT FILING DATE: 2000-04-14
;   PRIOR APPLICATION NUMBER: 60/241,755
;   PRIOR FILING DATE: 2000-10-20
;   PRIOR APPLICATION NUMBER: 60/237,768
;   PRIOR FILING DATE: 2000-10-03
;   PRIOR APPLICATION NUMBER: 60/231,498
;   PRIOR FILING DATE: 2000-09-08
;   NUMBER OF SEQ ID NOS: 207012
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 26413
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-26413

Query Match   3.3%;   Score 105.6;   DB 3;   Length 601;
Best Local Similarity   55.9%;   Pred. No. 3.2e-18;
Matches 198; Conservative   1; Mismatches 155; Indels   0; Gaps   0;

Qy   803 AAGTACCAGAATTGAAGGAGAGAAAGCACAGTTTTGGAGGACATCTCCACCTATGGA 862
             || |||| |||||||||||||||||||||||||||
Db   354 AGTTTTCAGAACCTTCTGTGGTCAGAAAGACATTTGTGGACAACATGAAAATTATAAC 295

Qy   863 GATGCATTCCTCCTCCTGCCAGCATTTCCTATCGGGCCAACACACAGGCATCTCTTTTAAA 922
             |||||||||||||| |||||||||||||||||| |
Db   294 CACAGTTACATCTACATCCCTGCTTTTCTATGAAGACAGGAACAGGAGCCATCTTTGAGG 235

Qy   923 GTCTACCAAACACTCAAAGAGTCAAAATGAGGCAAAAGGTTCTTCTTCCATCCCAGG 982
             |||||||||||||||||| |||
Db   234 GTTTATTATACACTGTCAGATGGTGCCAATCAAACAGTGCTGTTTGCCAACCCCAAC 175

Qy   983 TACCTGAGACACCTGCTCTTTTCTGGAGAACTAAAGGGGTGACTGCATACCGCTTGTCC 1042
             ||||||||||||||||||||||| |||||||||||
Db   174 TTTCTGCGTAGCATTGGAAAGTTCTGGAAAAGTAGAGGAATCCATGCCAAGCGCCTGTCC 115

Qy   1043 ACAGGCTTGATGATTGCAAGTGCGCGTGTGGAACTGTGTGAAAACGTGAGCTCTACGGA 1102
              ||||||||||||||| |||||| |||||||||||||
Db   114 ACAGGACTTTTCTGTGAGCGCGGCTCTGTGGGTCTCTGTGAAGAGGTGGCATCTATGGC 55

Qy   1103 TTCTGGCCTTTCTCTAAGACTATCGAAGACTATCAGTCCACCACTACTAT 1156
              ||||||||||||||| | |||||||
Db   54 TTCTGGCCCTTCTCGTGAATACATGAGCAGCTGGACCCATGTGAGACCTCAC 1

RESULT 7
US-09-949-016-72792/c
; Sequence 72792, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;   APPLICANT: VENTER, J. Craig et al.
;   TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;   TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;   FILE REFERENCE: CL001307
;   CURRENT APPLICATION NUMBER: US/09/949,016
;   CURRENT FILING DATE: 2000-04-14
;   PRIOR APPLICATION NUMBER: 60/241,755
;   PRIOR FILING DATE: 2000-10-20
;   PRIOR APPLICATION NUMBER: 60/237,768
;   PRIOR FILING DATE: 2000-10-03
;   PRIOR APPLICATION NUMBER: 60/231,498

; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 72792
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-72792

Query Match   3.3%;   Score 105.6; DB 3; Length 601;
Best Local Similarity   55.9%;   Pred. No. 3.2e-18;
Matches 198; Conservative   1; Mismatches 155; Indels   0; Gaps   0;

Qy   803 AAGTACCAGAGAATTTGAAGGAGGAAAGCACAGTTTTTGGAGGACATCTCCACCTATGGA 862
Db   354 AGGTTTCAGAAACCTTCTGTGGTCCAGAAAGACATTTGTGACAACATGAAAATYTATAAC 295

Qy   863 GATGCAATTCCTCCTCCTGCCAGCATTTCCTATGCGGGCCAACACAGGCATCTCTTTTAAA 922
Db   294 CACAGTTACATCTACATGCCTGCCTTTTCTATGAAGACAGGAACAGAGCCATCTTTGAGG 235

Qy   923 GTCTACCAAACACTCAAAGAGTCAAAAATGAGGCAAAAGGTTCTTCTTCCATCCCAGG 982
Db   234 GTTTATTATTATACACTGTCAGATGTTGGTGCCAATCAAACAGTGCGTGTTTGCCAACCCCAAC 175

Qy   983 TACCTGAGACACCTCGCTCTTTCTGAGAACTAAAGGGGTGACTGCATACCGGCTTGTCC 1042
Db   174 TTTCTGCGTAGCATTGGAAAGTTCTGGAAAGTAGAGGAATCCATGCCAAGCCCTGTCC 115

Qy 1043 ACAGGCTTGATGATTGCAAGTGTCGCTGTGTGGAAACGTGAAGTCTACGGA 1102
Db   114 ACAGGACTTTTTCTGGTGAGCGCAGCTCGGGTCTCTGTGAAGAGGTGGCCATCTATGGC 55

Qy 1103 TTCTGGCCTTTCTCTAAGACTATCGAAGACACACCCCACTCAGTCACCACTACTAT 1156
Db   54 TTCTGGCCCCTTCTCTGTGAATATGCATGAGCAGCAGCCCATCAGCCACCACTACTAT 1

RESULT 8
US-09-949-016-26414/c
; Sequence 26414, Application US/0994016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 26414
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-26414

Query Match   3.2%;   Score 100.2; DB 3; Length 601;
Best Local Similarity   61.1%;   Pred. No. 9.3e-17;
Matches 162; Conservative   0; Mismatches 103; Indels   0; Gaps   0;

Qy 1010 AGAACTAAAGGGGTGACTGCATACCGGCTTGTCCACAGGCTTGATGATTGCAAGTGTCGCT 1069
Db   601 AAAAGTAGAGGAATCCATGCCAAGCCCTGTCCACAGGACTTTTTCTGGTGAGCGCAGCT 542

Qy 1070 GTGGAACTGTGTGAAAACGTGAAGCTCTACGGATTCTGGCCTTTCTCTAAGACTATCGAA 1129

Db   541 CTGGGTCTCTGTGAAGAGGTGGCCATCTATGGCTTCTGGCCCCTTCTCTGTGAATATGCAT 482

Qy 1130 GACACCCCACTCAGTCACCACTACTATGATAACATGTTACCTAAGCATGGTTTCCACCAG 1189
Db   481 GAGCAGCCCATCAGCCACCACTACTATGACAACGTCTTACCCTTTTCTGGCTTCCATGCC 422

Qy 1190 ATGCCTAAAGAATACAGCCAAAGCTCCAGTCCATATGAGAGGAATCCTCAAACTGCAA 1249
Db   421 ATGCCCGAGGAATTTCTCCAACTTGGTATCTTCATAAAATCGGTGCACTGAGAATGCAG 362

Qy 1250 TTCAGCAAATGTGAAACGGCTTAAC 1274
Db   361 CTGGACCCATGTGAAGATACCTGAC 337

RESULT 9
US-09-949-016-72793/c
; Sequence 72793, Application US/0994016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 72793
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-72793

Query Match   3.2%;   Score 100.2; DB 3; Length 601;
Best Local Similarity   61.1%;   Pred. No. 9.3e-17;
Matches 162; Conservative   0; Mismatches 103; Indels   0; Gaps   0;

Qy 1010 AGAACTAAAGGGGTGACTGCATACCGGCTTGTCCACAGGCTTGATGATTGCAAGTGTCGCT 1069
Db   601 AAAAGTAGAGGAATCCATGCCAAGCCCTGTCCACAGGACTTTTTCTGGTGAGCGCAGCT 542

Qy 1070 GTGGAACTGTGTGAAAACGTGAAGCTCTACGGATTCTGGCCTTTCTCTAAGACTATCGAA 1129
Db   541 CTGGGTCTCTGTGAAGAGGTGGCCATCTATGGCTTCTGGCCCCTTCTCTGTGAATATGCAT 482

Qy 1130 GACACCCCACTCAGTCACCACTACTATGATAACATGTTACCTAAGCATGGTTTCCACCAG 1189
Db   481 GAGCAGCCCATCAGCCACCACTACTATGACAACGTCTTACCCTTTTCTGGCTTCCATGCC 422

Qy 1190 ATGCCTAAAGAATACAGCCAAAGCTCCAGTCCATATGAGAGGAATCCTCAAACTGCAA 1249
Db   421 ATGCCCGAGGAATTTCTCCAACTTGGTATCTTCATAAAATCGGTGCACTGAGAATGCAG 362

Qy 1250 TTCAGCAAATGTGAAACGGCTTAAC 1274
Db   361 CTGGACCCATGTGAAGATACCTGAC 337

RESULT 10
US-09-949-016-26415/c
; Sequence 26415, Application US/0994016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

```
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 26415
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-26415

Query Match        3.2%;  Score 99.8; DB 3;  Length 601;
Best Local Similarity  61.2%;  Pred. No. 1.2e-16;
Matches  161;  Conservative  0;  Mismatches  102;  Indels  0;  Gaps  0;

Qy   1012 AACTAAAGGGGTGACTGACTACCGCTTGTCCACAGGCTGTGATGATTGCAAGTGTCGCTGT 1071
          ||  || |||  |||||||||||||||  ||||||||||||||||||||||||||||||||
Db    600 AAGTAGAGGAATCATGCCAAGCGCCTGTCCACAGGACTTTTTCTGGTGAGCGCAGCTCT 541

Qy   1072 GGAACTGTGTGAAAACGTGAAGCTCTACGGATTCGGCCTTTCTCTAAGACTATCGAAGA 1131
          || ||||||||||||| ||||| |||||||||||  ||||  |||||||||||||||||
Db    540 GGGTCTCTGTGAAGAGGTGGCCATCTATGCGCCTTCTCGGCCCTTCTCTGTGAATATGCATGA 481

Qy   1132 CACCCCACTCAGTCACCACTACTATGATAACATGTTACCTAAGCACATGGTTTCCACCAGAT 1191
          ||||  |||| ||||  ||||||||||||||||||||  ||||| ||||| ||||||||||
Db    480 GCAGCCCATCAGCCACCACTACTA CTGACAACGTCTTACCCTTTTCTGGCTTCCATGCCAT 421

Qy   1192 GCCTAAAGAATACAGCCAAATGCCCAGCTCCATATGAGAGGAATCCTCAAACTGCAATT 1251
          |||  ||||| |||||||||||||||||| ||||||||||||||  ||||||||||||
Db    420 GCCCGAGGAATTTCTCCAACTCTGTATCTTCATAAAATCGGTGCACTGAGAATGCAGCT 361

Qy   1252 CAGCAAATGTGAAACGGCTTAAC 1274
          |||||||||||||| |||||||
Db    360 GGACCCATGTGAAGATACCTCAC 338
```

RESULT 11
US-09-949-016-72794/c
; Sequence 72794, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THERBOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 72794
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-72794

Query Match   3.2%; Score 99.8; DB 3; Length 601;
Best Local Similarity 61.2%; Pred. No. 1.2e-16;
Matches 161; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

Qy   1012 AACTAAAGGGGTGACTGACTACCGCTTGTCCACAGGCTGTGATGATTGCAAGTGTCGCTGT 1071
          || || ||| ||||||||||||||||||||||||||||||||||||||||||||||||
```

RESULT 12
US-08-626-994A-4
; Sequence 4, Application US/08626994A
; Patent No. 5798244
; GENERAL INFORMATION:
; APPLICANT: Shuichi TSUJI et al.
; TITLE OF INVENTION: Sia' 2,3Gala 1,4GlcNAc ' 2,8-
; TITLE OF INVENTION: SIALYLTRANSFERASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenblum & Bernstein, P.L.C.
; STREET: 1941 Roland Clarke Place
; CITY: Reston
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 20191
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 diskette
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Word Perfect 5.1+ (ASCII)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/626,994A
; FILING DATE: April 3, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 77469/1995
; FILING DATE: April 3, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Arnold Turk
; REGISTRATION NUMBER: 33,094
; REFERENCE/DOCKET NUMBER: P1495
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 716-1191
; TELEFAX: (703) 716-1180
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1048 bases
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
US-08-626-994A-4

Query Match   2.5%; Score 79.6; DB 2; Length 1048;
Best Local Similarity 50.2%; Pred. No. 5e-11;
Matches 283; Conservative 0; Mismatches 269; Indels 12; Gaps 3;

```
Qy    559 CATCCCCATTCGAGAGAACATTTCCAACATGTTTCCAGTGTCGCAGCCTTTTGTGGACTA 618
          |||  |||| |||||||||| || || |||  ||| |||||  ||||||||| |||||||
Db    301 CTTCTCGATTAGCAATAACTTCCGGTCCCCAGATGTGTCGCCCATTATGAATAA 360

Qy    619 TCCCTATAACCAGTGTGCAGTGCTTGGTAATGGGGGAATTCTCAACAGTCTCTCTGCGG 678
          ||  |||| |||||||||||||||||  ||||||  |||||||||||||||||||||||
Db    361 GCGTTATAATGTTTGTGCTGTGTGTGTTGGAAACAGTGGAATCTTGACGAGGAGTCAGTGTGG 420
```

```
;    LENGTH: 1048 bases
;    TYPE: nucleic acid
;    STRANDEDNESS: not relevant
;    TOPOLOGY: linear
US-08-957-742-4

Query Match             2.5%;  Score 79.6; DB 3; Length 1048;
Best Local Similarity  50.2%;  Pred. No. 5e-11;
Matches  283; Conservative    0; Mismatches  269; Indels   12; Gaps    3;

Qy    559 CATCCCCATTCGAGAGACACATT-----CCACATGTTCCAGTGTCGCAGCCTTTGTGGACTA 618
              ||||||  ||  ||| ||  |||        |||||||||||||||||||
Db    301 CTTCTCGATTAGCAATAACTTCCGTCCCTGCCCAGATGTGTCGCCATTATGAATAA 360

Qy    619 TCCCTATAACCAGTGTGCAGTGGTTGGTAATGGGGAATTCTCAACAAGTCTCTCTGCGG 678
              ||||||||||  |||||| |||||||||
Db    361 GCGTTATATGTTTGTGCGTGCGTGGAAACAGTGGAATCTTGACAGGAGTCAGTGTGG 420

Qy    679 AGCAGAAATTGATAAATCTGACTCGTCTTCAGTGTAACCTCCCCCAATCACAGGGAG 738
              ||| ||||| |||||||| ||| |
Db    421 ACAGAGAAATAGATAAATCAGATTTGTTTCTCGATGCAATTTTGCCCC---GACAGAGGC 477

Qy    739 CGCTAGTAAAGATGTTGGAAGCAAAACAAATCTTGTGACTGTCAATCCCAGCATTATAAC 798
                       ||||||||||||  |||||| | |
Db    478 TTTCCACAAGATGTTGGAAGGAAAACAACCTCACAACTTCTTCCACACTTCTTAGA 537

Qy    799 CCTGAAGTACCAGAATT-----GAAGGAGAAGAAAGCACAGTTTTTGGAGGACATCTC 852
              ||||||||||||          |
Db    538 GAAATATTACAACAATCTTTAACCATTCAGGACCGTAACAACTTCTTCCTCCTCAGTTTAAA 597

Qy    853 CACCTATGGAGATGCATTCCCTGCCCAGCATTTTCCTATCGGGCCAACACAGGCAT 912
              |||     |||| |   |
Db    598 AAAGCTTGATGGGGCCATACTTTGATCCCTGCATTTTCTTCCACACTTCTCTGCAACTGT 657

Qy    913 CTCTTTTAAAGTCTACCAAAACACTCAAAGAGTCAAAAATGAGGCAAAAGGTTCTCTTCTT 972
              ||       |||   || |||   |
Db    658 AACGAGAACGCTAGTGGATTTTTGTTGAGCACAGAGGTCAGTTAAAGGTCCAGTTGGC 717

Qy    973 CCATCCCAGGTA---CCTGAGAGACCCTCGCTCTTTTCTGAGAGAACTAAAGGGGTGACTGC 1029
              ||||||| ||      |||||||
Db    718 TTGGCCTGGAAATATCATGCAACTGTCAACAGTACTGGAAAAACAAACACCTGTCACC 777

Qy   1030 ATACCGCTTGTCCACAGGCTTGATGATTGCAAGTGTCGCTGTGGAACTGTGTGAAAACGT 1089
              || |    |  |||||| |    |
Db    778 CAAACGACTGAGCACAGGTATCCTAATGTATACTTCTTGCATCTGCAATATGTGAAGAGAT 837

Qy   1090 GAAGCTCTACGGATTCTGGCCTTT 1113
              ||  ||||||| |||||||||||
Db    838 CCACTTGTACGGTTTCTGGCCCTT 861
```

```
RESULT 13
US-08-957-742-4
; Sequence 4, Application US/08957742
; Patent No. 6017743
;  GENERAL INFORMATION:
;    APPLICANT: Shuichi TSUJI et al.
;    TITLE OF INVENTION: Sia'2,3Gala 1,4GlcNAc ' 2,8-
;    TITLE OF INVENTION: SIALYLTRANSFERASE
;    NUMBER OF SEQUENCES: 4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: Greenblum & Bernstein, P.L.C.
;      STREET: 1941 Roland Clarke Place
;      CITY: Reston
;      STATE: Virginia
;      COUNTRY: U.S.A.
;      ZIP: 20191
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: 1.44 diskette
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: MS-DOS
;      SOFTWARE: Word Perfect 5.1+ (ASCII)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/08/957,742
;      FILING DATE:
;      CLASSIFICATION: 435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 08/626,994
;      FILING DATE: April 3, 1996
;      APPLICATION NUMBER: JP 77469/1995
;      FILING DATE: April 3, 1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME: Arnold Turk
;      REGISTRATION NUMBER: 33,094
;      REFERENCE/DOCKET NUMBER: P14595
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE: (703) 716-1191
;      TELEFAX: (703) 716-1180
; INFORMATION FOR SEQ ID NO: 4:
;    SEQUENCE CHARACTERISTICS:
```

```
RESULT 14
US-08-626-994A-2
; Sequence 2, Application US/08626994
; Patent No. 5798244
;  GENERAL INFORMATION:
;    APPLICANT: Shuichi TSUJI et al.
;    TITLE OF INVENTION: Sia'2,3Gala 1,4GlcNAc ' 2,8-
;    TITLE OF INVENTION: SIALYLTRANSFERASE
;    NUMBER OF SEQUENCES: 4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: Greenblum & Bernstein, P.L.C.
;      STREET: 1941 Roland Clarke Place
;      CITY: Reston
;      STATE: Virginia
;      COUNTRY: U.S.A.
;      ZIP: 20191
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: 1.44 diskette
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: MS-DOS
;      SOFTWARE: Word Perfect 5.1+ (ASCII)
;    CURRENT APPLICATION DATA:
```

```
          APPLICATION NUMBER:  US/08/626,994A
          FILING DATE:  April 3, 1996
          CLASSIFICATION:  435
     PRIOR APPLICATION DATA:
          APPLICATION NUMBER:  JP 77469/1995
          FILING DATE:  April 3, 1995
     ATTORNEY/AGENT INFORMATION:
          NAME:  Arnold Turk
          REGISTRATION NUMBER:  33,094
          REFERENCE/DOCKET NUMBER:  P14595
     TELECOMMUNICATION INFORMATION:
          TELEPHONE:  (703) 716-1191
          TELEFAX:  (703) 716-1180
INFORMATION FOR SEQ ID NO: 2:
     SEQUENCE CHARACTERISTICS:
          LENGTH:  1660 bases
          TYPE:  nucleic acid
          STRANDEDNESS:  not relevant
          TOPOLOGY:  linear
US-08-626-994A-2
```

```
     TITLE OF INVENTION:  SIALYLTRANSFERASE
     NUMBER OF SEQUENCES:  4
     CORRESPONDENCE ADDRESS:
          ADDRESSEE:  Greenblum & Bernstein, P.L.C.
          STREET:  1941 Roland Clarke Place
          CITY:  Reston
          STATE:  Virginia
          COUNTRY:  U.S.A.
          ZIP:  20191
     COMPUTER READABLE FORM:
          MEDIUM TYPE:  1.44 diskette
          COMPUTER:  IBM PC compatible
          OPERATING SYSTEM:  MS-DOS
          SOFTWARE:  Word Perfect 5.1+ (ASCII)
     CURRENT APPLICATION DATA:
          APPLICATION NUMBER:  US/08/957,742
          FILING DATE:
          CLASSIFICATION:  435
     PRIOR APPLICATION DATA:
          APPLICATION NUMBER:  08/626,994
          FILING DATE:  April 3, 1996
          APPLICATION NUMBER:  JP 77469/1995
          FILING DATE:  April 3, 1995
     ATTORNEY/AGENT INFORMATION:
          NAME:  Arnold Turk
US-08-957-742-2
```

```
Query Match           2.5%;   Score 79.6;  DB 3;   Length 1660;
Best Local Similarity 50.2%;  Pred. No. 6.4e-11;
Matches  283;  Conservative   0;  Mismatches 269;  Indels  12;  Gaps   3;

Qy    559 CATCCCCATTCGAGAGAACATTTTCCACATGTTTCCAGTGTCGCAGCCTTTTGTGGACTA 618
              |||  |   || |||||  | ||   | ||  ||||| | |||| ||| ||| ||  ||
Db    485 CTTCTCGATTAGCAATAACTTCCGGTCCCTGCTCCCAGATGTGTCGCCCATTATGAATAA 544

Qy    619 TCCCTATAACCAGTGTGCAGTGGTTGGTAATGGGGAATTCTCAACAAGTCTCTCTGCGG 678
              | ||| |  ||| ||| | ||| || || |||| | ||  |||| |||| |||| ||
Db    545 GCGTTATATATGTTTGTGCTGTGTGGTGGAAACAGTGGAATCTTGACAGGGAGTCAGTGTGG 604

Qy    679 AGCAGAAATTGATAAATCTGACTTCGTCTTCAGGTGTAACCTCCCCCAATCACAGGGAG 738
             | |||||| |||||||| || ||| | ||| ||   ||     |||||| || |||| |
Db    605 ACAAGAAATAGATAAATCAGATTTGTTTCTCGATGCAATTTTGCCCC---GACAGAGGC 661

Qy    739 CGCTAGTAAAGATGTTGGAAGCAAAACAAATCTTGTGACTGTCAATCCCAGCATTATAAC 798
              | |  |||||||||||||| |||||| |  || ||| |  |||||| | |||| ||| |
Db    662 TTTCCACAAAGATGTTGGAAGGAAAACAACCTCACACACCTTCAATCCGAGCATCTTAGA 721

Qy    799 CCTGAAGTACCAGAATT------GAAGGAGAAGAAAGCACAGTTTTTGGAGGACATCTC 852
              | || |||| | ||||       |  ||| | ||  || ||||||  |||| |||| |
Db    722 GAAATATTACAACAATCTTTAACCATTCAGGACCGTAACAACTTCTTCCTCCAGTTTAAA 781

Qy    853 CACCTATGGAGATGCATTCCTCCTCTGCCAGATTTTCTATCGGGCCAACACAGGCAT 912
             |  | |||   | || ||||  | || | |||||||||   | | |||||| || |
Db    782 AAAGCTTGATGGGGCCATACTTTGGATCCCTGCATTTTTCTTCCACACTTCTGCAACTGT 841

Qy    913 CTCTTTTAAAGTCTACCAAAACACTCAAAGAGTCAAAAATGAGGCAAAAGGTTCTTCTT 972
             | | | |||| |  |   |||| || |||||| |||  |  |||  |||||| || |||
Db    842 AACGAGAACGCTAGTGAGTT---TGTTGAGCACAGAGGTCAGTTAAAGGTCAGTTGGC 901

Qy    973 CCATCCAGGTA---CCTGAGACCCTCGCTCTTTTCTGGAGAACTAAGGGGTGACTGC 1029
              | ||| || |    ||||| |||| |||  | ||||| | |||||| | |||| |
Db    902 TTGGCCTGGAAATATCATGCAACTGTCAACAGGTACTGGAAAAACAAACACCTGTCACC 961

Qy   1030 ATACCGCTTGTCCACAGGCTTGATGATTGCAAGTGCTGCTGTGGAAACGT 1089
             | ||| ||| | |||| | || |   ||| ||| | ||  ||||  ||
Db    962 CAAACGACTGAGCACAGCACAGGAGTATCCTAATGTATACTTGTGCAATATGTGAAGAGAT 1021

Qy   1090 GAAGCTCTACGGATTCTGGCCCTT 1113
             ||| | |||||| |||||||||||
Db   1022 CCACTTGTACGGTTTCTGGCCCTT 1045
```

RESULT 15
US-08-957-742-2
; Sequence 2, Application US/0895T742
; Patent No. 6017743
; GENERAL INFORMATION:
;   APPLICANT: Shuichi TSUJI et al.
;   TITLE OF INVENTION: Sia' 2,3Gala 1,4G1cNAc ' 2,8-

```
Qy   1030 ATACCGCTTGTCCACAGGCTTGATGATTGCAAGTGTCGCTGTGGAACTGTGTGAAAACGT 1089
               | ||| || ||||||| | || ||||||||| || || | | ||||||| |
Db    962 CAAACGACTGAGCACAGGTATCCTAATGTATACTCTTGCATCTGCAATATGTGAAGAGAT 1021

Qy   1090 GAAGCTCTACGGATTCTGGCCTTT 1113
               | ||||| ||||||| ||
Db   1022 CCACTTGTACGGTTTCTGGCCCTT 1045
```

Search completed: May 31, 2006, 14:25:31
Job time : 572 secs